



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Δημοσίευση Διασυνδεδεμένων Νομικών Δεδομένων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΡΗΓΙΝΑ-ΜΑΡΙΑ ΓΕΩΡΓΑΝΑ

Επιβλέπων : Παναγιώτης Τσανάκας
Καθηγητής Ε.Μ.Π.

Αθήνα, Μάρτιος 2024



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Δημοσίευση Διασυνδεδεμένων Νομικών Δεδομένων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΡΗΓΙΝΑ-ΜΑΡΙΑ ΓΕΩΡΓΑΝΑ

Επιβλέπων : Παναγιώτης Τσανάκας
ΚΑΘΗΓΗΤΗΣ Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 29^η Μαρτίου 2024.

.....
Παναγιώτης Τσανάκας
Καθηγητής Ε.Μ.Π.

.....
Ευγενία Τζαννίνη
Επίκουρη Καθηγήτρια Ε.Μ.Π.

.....
Ανδρέας-Γεώργιος Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.

Αθήνα, Μάρτιος 2024

.....

ΡΗΓΙΝΑ-ΜΑΡΙΑ ΓΕΩΡΓΑΝΑ

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Ρηγίνα-Μαρία Γεωργάνα, 2024.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Η ψηφιακή δημοσίευση της Ελληνικής Νομοθεσίας αντιμετωπίζει σημαντικές προκλήσεις όσον αφορά την προσβασιμότητα, την οργάνωση και την εμπειρία των χρηστών, υπογραμμίζοντας την επείγουσα ανάγκη για μια γενική αναθεώρηση. Καθώς οι νομικές πληροφορίες είναι μείζονος σημασίας για τη συμμετοχή των πολιτών, η αποτελεσματικότητα της διαδικτυακής διάδοσής τους δεν μπορεί να υποτιμηθεί. Η παρούσα Διπλωματική Εργασία αξιοποιεί τις δυνατότητες του Σημασιολογικού Ιστού για να επαναπροσδιορίσει τον τρόπο με τον οποίο τα ελληνικά νομικά κείμενα δημοσιεύονται διαδικτυο.

Η εργασία αυτή αντλεί έμπνευση από την οντολογία/πρότυπο ELI της Ευρωπαϊκής Ένωσης, η οποία αποτελεί ένα ενοποιημένο και τυποποιημένο μοντέλο δημοσίευσης νομικών πληροφοριών για τα κράτη-μέλη της. Προσαρμόζει σχολαστικά και επεκτείνει το πρότυπο ELI έτσι ώστε να μπορεί να ανταποκριθεί στην περίπλοκη φύση της Ελληνικής Νομοθεσίας, δημιουργώντας την οντολογία ELIgr. Αυτή η νέα οντολογία ανοίγει το δρόμο για τη μετατροπή νομικών εγγράφων απαρχαιωμένης μορφής σε σημασιολογικά έγγραφα RDF, ένα άλμα προς την αξιοποίηση της δύναμης των Διασυνδεδεμένων Δεδομένων για τη δημοσίευση της Ελληνικής Νομοθεσίας.

Παράλληλα με τη θεωρητική ανάπτυξη του ELIgr, δίνεται σημασία και στην πρακτική εφαρμογή του: τη μετατροπή των νομικών κειμένων σε αναζητήσιμα διασυνδεδεμένα σημασιολογικά έγγραφα καθιστώντας τα προσβάσιμα μέσω ενός διακομιστή. Η πρόσβαση αυτή επιτυγχάνεται με τη βοήθεια SPARQL ερωτημάτων, μέσω των οποίων παρουσιάζουμε σενάρια χρήσης. Επιπλέον, στο τέλος της παρούσας εργασίας, θέτουμε τις βάσεις για μελλοντικές βελτιώσεις, στοχεύοντας σε ένα ψηφιακό οικοσύστημα όπου η Ελληνική Νομοθεσία δεν παρουσιάζεται απλά, αλλά δημοσιεύεται και χρησιμοποιείται με τρόπο σύγχρονο και αποτελεσματικό.

Λέξεις Κλειδιά: Σημασιολογικός Ιστός, Διασυνδεδεμένα Δεδομένα, Ελληνική Νομοθεσία, ELI

Abstract

The digital publication of Greek Legislation faces significant challenges in terms of accessibility, organization and user experience, highlighting the urgent need for an overhaul. As legal information is of major importance for citizen participation, the effectiveness of its online dissemination cannot be underestimated. This Thesis utilizes the potential of the Semantic Web to redefine the way in which greek legal texts are published online.

This work draws inspiration from the European Union's ELI ontology/standard, which is a unified and standardized model for publishing legal information for its member-states. It meticulously adapts and extends the ELI standard so that it can respond to the complex nature of Greek Legislation, creating the ELIgr ontology. This new ontology paves the way for the conversion of obsolete legal documents into semantic RDF documents, a leap towards harnessing the power of Linked Data for the publication of Greek Legislation.

Along with the theoretical development of ELIgr, importance is also given to its practical application: the transformation of legal texts into searchable interconnected semantic documents by making them accessible through a server. This access is achieved with the help of SPARQL queries, through which we present usage scenarios. In addition, at the end of this paper, we lay the foundations for future improvements, aiming for a digital ecosystem where Greek Legislation is not simply presented, but published and used in a modern and efficient way.

Keywords: Semantic Web, Linked Data, Greek Legislation, ELI

*στη μνήμη του παππού μου,
Γκερέκου Χρήστου
γνωστού ως
«Χαρίτων ο Γλαφυρός»*

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου, κύριο Παναγιώτη Τσανάκα, για την εμπιστοσύνη που μου έδειξε και για την ευκαιρία που μου έδωσε να εκπονήσω τη Διπλωματική μου Εργασία πάνω σε έναν τόσο σημαντικό και ενδιαφέροντα τομέα. Ακόμα, θα ήθελα να εκφράσω την απεριόριστη ευγνωμοσύνη μου στο μέλος Ε.ΔΙ.Π. κύριο Μάριο Κόνιαρη για την υπομονή του, την πολύτιμη βοήθειά του και τις συμβουλές που μου προσέφερε κατά τη διάρκεια εκπόνησης της παρούσας Διπλωματικής Εργασίας.

Θα ήθελα, επίσης, να ευχαριστήσω θερμά τους γονείς και τους παππούδες μου για τη στήριξη κάθε είδους που μου έχουν προσφέρει καθ' όλη τη διάρκεια της φοίτησής μου στο Εθνικό Μετσόβιο Πολυτεχνείο και, γενικότερα, της ζωής μου. Τέλος, θα ήθελα να ευχαριστήσω ιδιαίτερα τον σύντροφό μου, Γρηγόρη, καθώς και τους αγαπημένους φίλους μου, Δήμητρα, Μάριο, Κωνσταντίνα, Κωστή, Γιάννη και Νικολέτα για την αγάπη και την ηθική συμπαράσταση που μου προσέφεραν απλόχερα όποτε τη χρειάστηκα.

Γεωργάνα Ρηγίνα-Μαρία

Πίνακας Περιεχομένων

1 ΕΙΣΑΓΩΓΗ.....	1
1.1 Περιγραφή του Προβλήματος.....	1
1.2 Διεθνής Ανασκόπηση.....	2
1.3 Η Ελληνική πραγματικότητα.....	3
1.4 Αντικείμενο και Στόχοι της Παρούσας Εργασίας.....	5
1.5 Σχετικές Εργασίες.....	6
1.6 Οργάνωση Κειμένου.....	8
2 ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ.....	9
2.1 Σημασιολογικός Ιστός (Semantic Web) και Διασυνδεδεμένα Δεδομένα (Linked Data)....	9
2.1.1 Σημασιολογικός Ιστός (Semantic Web).....	9
2.1.2 Διασυνδεδεμένα Δεδομένα (Linked Data).....	11
2.1.3 Οντολογίες (Ontologies).....	14
2.1.4 Λεξιλόγια (Vocabularies).....	16
2.1.5 Χώροι Ονομάτων (Namespaces).....	16
2.1.6 Μεταδεδομένα (Metadata)	17
2.2 Τεχνολογίες στο Σημασιολογικό Ιστό.....	18
2.2.1 Το πρότυπο RDF.....	18
2.2.2 Ενιαίος Ταυτοποιητής Πόρων (Uniform Resource Identifier - URI) και Παγκόσμιος Ταυτοποιητής Πόρων (International Resource Identifier - IRI).....	20
2.2.3 Γλώσσες οντολογίας (Ontology Languages).....	21
2.2.4 Extensible Markup Language (XML).....	24
2.2.5 Γλώσσες Ερωτημάτων στο Σημασιολογικό Ιστό.....	25
2.2.6 Εργαλεία Δημοσίευσης Διασυνδεδεμένων Δεδομένων.....	26
2.2.7 Άλλες Τεχνολογίες στο Σημασιολογικό Ιστό.....	28
2.3 Το Πρότυπο ELI.....	29
2.3.1 Τεχνικά Στοιχεία.....	29
2.3.2 Επεκτάσεις και Εργαλεία της Οντολογίας ELI.....	33
3 ΑΝΑΠΑΡΑΣΤΑΣΗ ΕΛΛΗΝΙΚΗΣ ΝΟΜΟΘΕΣΙΑΣ.....	34
3.1 Μοντέλο Διασυνδεδεμένων Δεδομένων Ελληνικής Νομοθεσίας.....	34
3.2 Παράδειγμα Εφαρμογής.....	41
4 ΣΗΜΑΣΙΟΛΟΓΙΚΑ ΕΓΓΡΑΦΑ ΣΤΟ ELIgr.....	45
4.1 Διαδικασία Αυτόματης Μετατροπής των Σημασιολογικών Εγγράφων από το Πρότυπο LegalDocML στο Πρότυπο RDF.....	45
4.2 Περιγραφή της Δομής των LegalDocML Σημασιολογικών Εγγράφων.....	46
4.3 Περιγραφή της Δομής και της Λειτουργίας του Μετασχηματισμού XSLT.....	50
5 ΣΕΝΑΡΙΑ ΧΡΗΣΗΣ.....	54
5.1 Συγκεντρωτική Λίστα Είδους Νομοθετήματος έως συγκεκριμένη Ημερομηνία.....	54
5.2 Συγκεντρωτική Λίστα Είδους Νομοθετήματος συγκεκριμένου Έτους.....	55
5.3 Συγκεντρωτική Λίστα Άρθρων συγκεκριμένης Πηγής Νομοθεσίας.....	56
5.4 Συγκεντρωτική Λίστα Άρθρων συγκεκριμένης Πηγής Νομοθεσίας και αριθμός των Παραγράφων τους.....	57
5.5 Συγκεντρωτική Λίστα Αναφορών σε άλλες Πηγές Νομοθεσίας συγκεκριμένης Πηγής Νομοθεσίας.....	58
5.6 Συγκεντρωτική Λίστα Πηγών Νομοθεσίας που περιέχουν Αναφορές σε άλλες Πηγές	

Νομοθεσίας και Αριθμός των Αναφορών αυτών σε Φθίνουσα Σειρά.....	59
6 ΕΠΙΛΟΓΟΣ.....	.65
6.1 Σύνοψη.....	.65
6.2 Συμπεράσματα και Παρατηρήσεις.....	.66
6.3 Συνεισφορά και Προτάσεις για το Μέλλον.....	.66
7 ΒΙΒΛΙΟΓΡΑΦΙΑ.....	.69

Πίνακας Περιεχομένων Εικόνων

Εικόνα 1: Στοίβα του Σημασιολογικού Ιστού.....	10
Εικόνα 2: Το LOD Cloud του έτους 2014.....	13
Εικόνα 3: Το LOD Cloud του έτους 2022.....	13
Εικόνα 4: Παράδειγμα Οντολογίας στο Νομικό Πεδίο.....	15
Εικόνα 5: Παράδειγμα RDF Γράφου για το Νόμο 3918/2011.....	19
Εικόνα 6: Συντακτικό Διάγραμμα των URI.....	20
Εικόνα 7: Η βασική μορφή του πρότυπου ELI.....	32
Εικόνα 8: Οι βασικές κλάσεις της οντολογίας ELI.....	35
Εικόνα 9: Οι κλάσεις της οντολογίας ELI και οι μεταξύ τους σχέσεις.....	36
Εικόνα 10: Τα Είδη των Νομικών Πράξεων στο ELIgr.....	37
Εικόνα 11: Τα Είδη των Υποδιαιρέσεων Νομικών Πράξεων στο ELIgr.....	38
Εικόνα 12: Οι προσθήκες στην κλάση Agent στο ELIgr.....	39
Εικόνα 13: Βασικές Πληροφορίες της Πηγής Νομοθεσίας στο RDF αρχείο.....	41
Εικόνα 14: Περιεχόμενα Πηγής Νομοθεσίας (Λίστα Κεφαλαίων) στο RDF αρχείο.....	42
Εικόνα 15: Περιεχόμενα Πηγής Νομοθεσίας (Λίστα Άρθρων Κεφαλαίου) στο RDF αρχείο	42
Εικόνα 16: Περιεχόμενα Πηγής Νομοθεσίας (Λίστα Παραγράφων Άρθρου) στο RDF αρχείο	42
Εικόνα 17: Αναφορές (Citations) σε άλλες Πηγές Νομοθεσίας στο RDF αρχείο.....	43
Εικόνα 18: Περιεχόμενο Εξωφύλλου Πηγής Νομοθεσίας στο RDF αρχείο.....	43
Εικόνα 19: Περιεχόμενο Εισαγωγής Πηγής Νομοθεσίας στο RDF αρχείο.....	43
Εικόνα 20: Περιεχόμενο Άρθρου Πηγής Νομοθεσίας στο RDF αρχείο.....	44
Εικόνα 21: Περιεχόμενο Επιλόγου Πηγής Νομοθεσίας στο RDF αρχείο.....	44
Εικόνα 22: Διάγραμμα Ροής Διαδικασίας Μετατροπής των LegalDocML σημασιολογικών εγγράφων σε RDF.....	46
Εικόνα 23: Μεταδεδομένα LegalDocML Αρχείου (μέρος 2ο).....	47
Εικόνα 24: Μεταδεδομένα LegalDocML Αρχείου (μέρος 1ο).....	47
Εικόνα 25: Εξώφυλλο και Δομή της Πηγής Νομοθεσίας στο LegalDocML Αρχείο (μέρος 1ο).....	48
Εικόνα 26: Εξώφυλλο και Δομή της Πηγής Νομοθεσίας στο LegalDocML Αρχείο (μέρος 2ο).....	48
Εικόνα 27: Εξώφυλλο και Δομή της Πηγής Νομοθεσίας στο LegalDocML Αρχείο (μέρος 3ο).....	48
Εικόνα 28: Εισαγωγικό Κείμενο της Πηγής Νομοθεσίας στο LegalDocML Αρχείο.....	48
Εικόνα 29: Περιεχόμενο κάθε υποενότητας της Πηγής Νομοθεσίας στο LegalDocML Αρχείο.....	49
Εικόνα 30: Κείμενο Επιλόγου της Πηγής Νομοθεσίας στο LegalDocML Αρχείο.....	49
Εικόνα 31: Εγγραφή Μεταδεδομένων στα RDF Σημασιολογικά Έγγραφα.....	50
Εικόνα 32: Εγγραφή των Τμημάτων της Πηγής Νομοθεσίας στα RDF Σημασιολογικά Έγγραφα.....	51
Εικόνα 33: Εγγραφή των Αναφορών της Πηγής Νομοθεσίας σε άλλες Πηγές Νομοθεσίας στα RDF Σημασιολογικά Έγγραφα.....	51
Εικόνα 34: Εγγραφή του Περιεχομένου Άρθρων και Παραγράφων στα RDF Σημασιολογικά Έγγραφα.....	52
Εικόνα 35: SPARQL Ερώτημα Q1.....	55
Εικόνα 36: SPARQL Ερώτημα Q2.....	56
Εικόνα 37: SPARQL Ερώτημα Q3.....	57
Εικόνα 38: SPARQL Ερώτημα Q4.....	58

Εικόνα 39: SPARQL Ερώτημα Q5.....	59
Εικόνα 40: SPARQL Ερώτημα Q6.....	59

Πίνακας Περιεχομένων Πινάκων

Πίνακας 1: Συστατικά Στοιχεία των URI του προτύπου ELI.....	30
Πίνακας 2: Συντομογραφίες Τύπων Νομικών Εγγράφων στο ELIgr.....	40
Πίνακας 3: Αποτελέσματα SPARQL Ερωτήματος Q1.....	55
Πίνακας 4: Αποτελέσματα SPARQL Ερωτήματος Q2.....	56
Πίνακας 5: Αποτελέσματα SPARQL Ερωτήματος Q3.....	57
Πίνακας 6: Αποτελέσματα SPARQL Ερωτήματος Q4.....	58
Πίνακας 7: Αποτελέσματα SPARQL Ερωτήματος Q5.....	59
Πίνακας 8: Αποτελέσματα SPARQL Ερωτήματος Q6.....	64

1

Εισαγωγή

1.1 Περιγραφή του προβλήματος

Ένα από τα σημαντικότερα πλεονεκτήματα που μας δίνει η σύγχρονη ψηφιακή εποχή είναι η ευκολότερη και πιο ισότιμη πρόσβαση στην πληροφορία. Σε αντίθεση με παλαιότερες εποχές, καθένας από εμάς μπορεί με απλές ενέργειες να βρει κάθε πληροφορία πάνω στο θέμα που αναζητεί. Δεδομένου όμως του τεράστιου όγκου της πληροφορίας που δημοσιεύεται καθημερινά στο διαδίκτυο, δημιουργείται ένα νέο πρόβλημα, αυτό της σωστής ταξινόμησης και σύνδεσης των πληροφοριών. Ακριβώς επειδή ο όγκος της πληροφορίας είναι τόσο μεγάλος, υπάρχει επιτακτική ανάγκη η πληροφορία αυτή να γίνει διαθέσιμη με τρόπο απλούστερο και πιο κατανοητό.

Ένα πρώτο βήμα προς αυτή την κατεύθυνση ήταν η δημιουργία του Παγκοσμίου Ιστού (World Wide Web) και η ανάπτυξη βάσεων δεδομένων (databases). Μέσω του Παγκοσμίου Ιστού, η πρόσβαση στην πληροφορία έγινε δυνατή για όλους μας, ενώ κάθε οργανισμός ή υπηρεσία, δημόσια ή ιδιωτική, μπορεί να ταξινομήσει τα δεδομένα της σε κατάλληλες βάσεις δεδομένων προσαρμοσμένες στις ανάγκες της. Με αυτόν τον τρόπο, η αποθήκευση της πληροφορίας γίνεται πιο οργανωμένα και με τάξη, κάνοντας, ως εκ τούτου, την εύρεσή της πολύ ευκολότερη. Τι συμβαίνει, όμως, όταν χρειαζόμαστε πληροφορίες από δύο, τρεις, ή και παραπάνω βάσεις δεδομένων; Πώς μπορούμε με ευκολία να συγκρίνουμε, να συνδέσουμε, να επεξεργαστούμε και να αξιοποιήσουμε πληροφορίες από πάνω από μια πηγές, ειδικά όταν η προσπάθεια αυτή γίνεται πάνω σε μεγάλη ποσότητα πληροφορίας την οποία μόνο ένας υπολογιστής μπορεί να διαχειριστεί;

Το επόμενο βήμα προς αυτήν την κατεύθυνση αποτελεί ο Σημασιολογικός Ιστός (Semantic Web) και τα Διασυνδεδεμένα Δεδομένα (Linked Data). Η σύλληψη της ιδέας του Σημασιολογικού Ιστού οφείλεται στον Tim Berners-Lee, εφευρέτη του Παγκοσμίου Ιστού και διευθυντή της Κοινοπραξίας του Παγκοσμίου Ιστού (World Wide Web Consortium ή W3C), οργάνωσης με αντικείμενο την ανάπτυξη διεθνών προτύπων για τον Παγκόσμιο Ιστό. Ο Σημασιολογικός Ιστός είναι μια επέκταση του Παγκοσμίου Ιστού, όπου η πληροφορία θα είναι δομημένη με τέτοιο τρόπο ώστε να μπορεί ένας υπολογιστής να τη διαβάσει, να την επεξεργαστεί και να εκτελέσει σύνθετες εργασίες με βάση αυτή για την εξυπηρέτηση του εκάστοτε χρήστη. Σε αντίθεση με το σήμερα, όπου το μεγαλύτερο ποσοστό της πληροφορίας στο ίντερνετ είναι σχεδιασμένο για

«κατανάλωση» από ανθρώπους, ο Σημασιολογικός Ιστός προϋποθέτει πως όλες οι πληροφορίες είναι δημοσιευμένες με βάση κατάλληλες και συγκεκριμένες πρακτικές, ώστε να μπορούν να συνδεθούν σε ένα «δίκτυο πληροφοριών» και να χρησιμοποιηθούν από έναν υπολογιστή.

Ανάμεσα σε όλους τους τομείς στους οποίους μπορεί να μας εξυπηρετήσει ο Σημασιολογικός Ιστός, ιδιαίτερο ενδιαφέρον παρουσιάζει η νομοθεσία. Μελετώντας και συγκρίνοντας όλους τους νόμους ενός ή και περισσότερων κρατών μέσα στη ροή της ιστορίας, μπορούν να εξαχθούν πολύ χρήσιμα συμπεράσματα για την εξέλιξη, τη νοοτροπία, την κουλτούρα και τις αντιλήψεις κάθε λαού, αλλά και της ανθρωπότητας ως σύνολο ανά τους αιώνες. Επιπλέον, είναι ιδιαίτερος σημαντικό ο κάθε πολίτης να είναι ενημερωμένος για τη νομοθεσία της χώρας του και να μπορεί να τη συγκρίνει με την αντίστοιχη νομοθεσία άλλων κρατών. Η γνώση αυτή θα τον καταστήσει ενεργό πολίτη, ο οποίος γνωρίζει καλά τα δικαιώματα και τις υποχρεώσεις του, με αποτέλεσμα να σχεδιάζει με πιο αποδοτικό τρόπο τη ζωή του και να μπορεί να έχει μια λογική βάση στις αξιώσεις του για αλλαγές από τις εκάστοτε κυβερνήσεις.

Για να επιτευχθούν οι παραπάνω στόχοι, αρκετά κράτη έχουν ήδη αρχίσει να χρησιμοποιούν τις τεχνολογίες του Σημασιολογικού Ιστού κατά τη δημοσίευση των νομικών πληροφοριών τους. Η Ελλάδα, ως ένα σύγχρονο Ευρωπαϊκό κράτος, οφείλει να μη μείνει πίσω σε αυτόν τον τομέα. Ωστόσο, παρότι συναντούμε ορισμένες εργασίες και ιδιωτικές πρωτοβουλίες που κινούνται προς αυτή τη κατεύθυνση, η απουσία ενός επίσημου κεντρικού συστήματος διαχείρισης και διανομής των νομικών πληροφοριών του ελληνικού κράτους σε μορφή Διασυνδεδεμένων Δεδομένων που αξιοποιεί τις δυνατότητες του Σημασιολογικού Ιστού είναι αισθητή.

1.2 Διεθνής Ανασκόπηση

Τα τελευταία χρόνια, τα νομικά δεδομένα είναι περισσότερο διαθέσιμα στο ευρύ κοινό σε σχέση με οποιαδήποτε άλλη χρονική περίοδο. Οι περισσότερες χώρες, εκμεταλλευόμενες τη ραγδαία ανάπτυξη των βάσεων δεδομένων και των εργαλείων σχεδιασμού, καθώς και την διαδεδομένη χρήση του ίντερνετ, έχουν δημιουργήσει τα δικά τους συστήματα και ιστοσελίδες ώστε να παρέχουν στους πολίτες τους τη νομοθεσία της χώρας τους. Ωστόσο, τα συστήματα αυτά δεν είναι πάντα λειτουργικά. Συχνό είναι το φαινόμενο οι ιστοσελίδες να είναι σχεδιασμένες με βάση αρκετά παλαιότερα πρότυπα, να παραλείπεται η τακτική ενημέρωση και συντήρησή τους, να καλούνται να εξυπηρετήσουν περισσότερους χρήστες από αυτούς για τους οποίους είχαν σχεδιαστεί ή να είναι αρκετά περίπλοκες για το μέσο χρήστη. Επιπλέον, τις περισσότερες φορές, η νομοθεσία διατίθεται αποκλειστικά σε πιο «απαρχαιωμένες» μορφές αρχείων, όπως docx ή pdf, ενώ οι πιο σύγχρονες προσεγγίσεις περιλαμβάνουν τη νομοθεσία και σε μορφή html.

Ωστόσο, έχουν γίνει ορισμένες προσπάθειες για τη δημοσίευση των νομικών δεδομένων στον Σημασιολογικό Ιστό. Στις Η.Π.Α., σχεδόν όλη η νομοθεσία είναι διαθέσιμη, ανάμεσα σε άλλα, και σε μορφή σημασιολογικών εγγράφων (XML) μέσω της **ιστοσελίδας**¹ του Αμερικανικού Κυβερνητικού Εκδοτικού Γραφείου (US Government Publishing Office ή GPO), η οποία μάλιστα αντικατέστησε την παλαιότερη ιστοσελίδα του FDSys που υπήρχε για αυτόν τον σκοπό. Η Ευρωπαϊκή Ένωση διαθέτει το Ευρωπαϊκό Δίκαιο σε διάφορες γλώσσες και μορφές, μεταξύ αυτών και η σημασιολογική (XML), μέσω της ιστοσελίδας **EUR-Lex**². Επιπλέον, η υπηρεσία **N-Lex**³ της Ευρωπαϊκής Ένωσης αποτελεί μια προσπάθεια σύνδεσης των νομοθεσιών όλων των κρατών-μελών της, η οποία ωστόσο εξαρτάται από την πολιτική που έχει εφαρμόσει κάθε χώρα, με αποτέλεσμα άλλες χώρες να διαθέτουν τα νομικά δεδομένα τους σε πιο παλαιωμένες μορφές (όπως η Λετονία, η Κύπρος, η Ουγγαρία κ.ο.κ.) και άλλες να χρησιμοποιούν και το Σημασιολογικό Ιστό (όπως η Ισπανία, η Δανία, η Κροατία κ.ο.κ. Η Αυστρία διαθέτει τους νόμους της στο κοινό σε XML μορφή

1 <https://www.govinfo.gov/>

2 <https://eur-lex.europa.eu/homepage.html>

3 <https://n-lex.europa.eu/n-lex/index?lang=en>

μέσω της υπηρεσίας **RIS**⁴, ενώ στη Φινλανδία αναπτύχθηκε το Semantic Finlex Legislation. Τέλος, άξια αναφοράς είναι και η κυβερνητική **ιστοσελίδα**⁵ του Ηνωμένου Βασιλείου, η οποία εκμεταλλεύεται σε αρκετά μεγάλο, συγκριτικά, βαθμό τις δυνατότητες που προσφέρουν τα Διασυνδεδεμένα Δεδομένα, κάνοντας ευκολότερη την αλληλεπίδραση με τον χρήστη.

Επιπλέον, το 2012, η Ευρωπαϊκή Ένωση αποφάσισε τη θέσπιση του Αναγνωριστικού Ευρωπαϊκής Νομοθεσίας (European Legislation Identifier - ELI) [1]. Η Ευρωπαϊκή Ένωση ενθαρρύνει τα κράτη-μέλη της να δημοσιεύουν τη νομοθεσία τους στο διαδίκτυο με βάση το κοινό για όλους πρότυπο ELI, με στόχο την ευκολότερη και γρηγορότερη ανταλλαγή νομικών πληροφοριών μεταξύ τους. Με αυτόν τον τρόπο θα διευκολυνθεί η συνεργασία τους, καθώς θα υπάρχει ισχυρότερη διαλειτουργικότητα μεταξύ των πληροφοριακών συστημάτων των εθνικών και ευρωπαϊκών θεσμικών οργάνων.

1.3 Η Ελληνική Πραγματικότητα

Στην Ελλάδα, η νομοθεσία παρουσιάζει πολλά προβλήματα και ως προς την διαθεσιμότητά της στο κοινό, αλλά και ως προς τη δομή και το περιεχόμενο των νόμων. Τα φαινόμενα της πολυνομίας και της κακονομίας, παρότι έχουν αναγνωριστεί από τις αρχές του 1950, είναι διευρυμένα ακόμα και σήμερα. Υπάρχουν νόμοι οι οποίοι δεν έχουν εκσυγχρονιστεί ώστε να ακολουθούν τις εξελίξεις και να εξυπηρετούν τις ανάγκες της εποχής μας σε οικονομικό, κοινωνικό, αλλά και τεχνολογικό επίπεδο. Συχνά βλέπουμε περιπτώσεις στις οποίες νόμοι αντικρούουν ή επικαλύπτουν ο ένας τον άλλον και συναντούμε ασάφειες, αχρείαση περιπλοκότητα, λογοτεχνικό ύφος, λόγια γλώσσα και επαναληψιμότητα. Επιπλέον, η ελληνική νομοθεσία είναι, τις περισσότερες φορές, δημοσιευμένη στο διαδίκτυο σε μορφή απλού κειμένου χωρίς να γίνεται καμία αναφορά στο κατά πόσο το εκάστοτε νομικό κείμενο βρίσκεται σε ισχύ, χωρίς να υπάρχουν σελιδοδείκτες για τα διαφορετικά κεφάλαια ή τα σημαντικότερα τμήματα του νομικού κειμένου και χωρίς να υπάρχει κάποια περίληψή του ή θεματική του ταξινόμηση. Αν οι επιλογές αυτές ήταν διαθέσιμες, τα προβλήματα που αναφέρθηκαν θα μετριάζονταν εν μέρει, καθώς ο πολίτης θα είχε τη δυνατότητα να περιηγηθεί στο νομικό κείμενο με μεγαλύτερη ευκολία, προσπερνώντας περιττές πληροφορίες. Η έλλειψη αυτών των εργαλείων, ωστόσο, αφήνει τον πολίτη έρμαιο των παραπάνω προβλημάτων και δημιουργεί ένα νομικό περιβάλλον αρκετά δυσπρόσιτο προς αυτόν.

Τα προβλήματα που προκύπτουν από την κατάσταση αυτή είναι πολλά και αφορούν μεγάλο μέρος της κοινωνίας μας. Συχνά οι πολίτες δε γνωρίζουν τα δικαιώματά τους, τις υποχρεώσεις τους και τις κατάλληλες ενέργειες για τη ρύθμιση των προσωπικών τους υποθέσεων. Αυτό έχει ως αποτέλεσμα να δημιουργούνται πολίτες «δύο ταχυτήτων», φέρνοντας σε μειονεκτική θέση όσους έχουν χαμηλότερο επίπεδο μόρφωσης ή δεν έχουν τους πόρους ώστε να συμβουλευτούν κάποιον κατάλληλο δικηγόρο. Σε επίπεδο κράτους και επιχειρήσεων, πολλές φορές συναντούμε περιττή γραφειοκρατία σε τεράστιο βαθμό, με απότοκο να δεσμεύονται σημαντικοί πόροι σε διαχειριστικά έξοδα. Έτσι, οι επιχειρήσεις αποθαρρύνονται από το να επενδύσουν στη χώρα μας, με οδυνηρές συνέπειες στην οικονομία και την ανάπτυξή της. Το κράτος, από τη μεριά του, αναγκάζεται να σπαταλά χρήματα τα οποία θα μπορούσαν να επενδυθούν σε πιο άμεσες ανάγκες που εξυπηρετούν τους πολίτες, όπως το εκπαιδευτικό σύστημα και το σύστημα υγείας. Επιπλέον, το κλίμα αυτό δυσκολεύει την ύπαρξη διαφάνειας στις πράξεις της Δημόσιας Διοίκησης, με αποτέλεσμα η Ελλάδα να είναι πρώτη στην Ευρωπαϊκή Ένωση σε επίπεδα διαφθοράς σύμφωνα με τους πολίτες της, με βάση την πιο πρόσφατη **έρευνα της Ευρωπαϊκής Επιτροπής**⁶.

Για την βελτίωση της κατάστασης και τη διευκόλυνση κράτους, επιχειρήσεων και πολιτών, η Ελλάδα, από τις αρχές του 1950 αλλά κυρίως μετά την είσοδό της στην Ευρωπαϊκή Οικονομική Κοινότητα (ΕΟΚ) και στην Οικονομική Νομισματική Ένωση (ΟΝΕ) εισήγαγε την έννοια της «καλής νομοθέτησης». Σύμφωνα με

4 <https://www.ris.bka.gv.at/default.aspx>

5 <https://www.legislation.gov.uk/>

6 <https://europa.eu/eurobarometer/surveys/detail/2658>

το **Γραφείο Καλής Νομοθέτησης**⁷, η ύπαρξη του οποίου θεσπίστηκε από το **N. 4048/2012**⁸, ορίζεται ως καλή νομοθέτηση «η πολιτική διαμόρφωσης αρχών και μέσων για τη βελτίωση της ποιότητας των ρυθμίσεων και των διαδικασιών παραγωγής τους». Ανάμεσα στις αρμοδιότητες του Γραφείου Καλής Νομοθέτησης, κατά το **Π.Α. 98/2020**⁹, είναι η λήψη όλων των αναγκαίων μέτρων για την εφαρμογή των αρχών και εργαλείων της καλής νομοθέτησης και ο συντονισμός των δράσεων κωδικοποίησης και αναμόρφωσης της νομοθεσίας. Ο Ν. 4048/2012 θέσπιζε επίσης τους κανόνες της καλής νομοθέτησης, χωρίς όμως να δίνει έμφαση στις διαδικασίες οι οποίες είναι απαραίτητες για την τήρηση των κανόνων αυτών. Για τον λόγο αυτό, αντικαταστάθηκε από το **N. 4622/2019**¹⁰, ο οποίος εφοδιάζει την εκτελεστική λειτουργία με ένα διεξοδικό εγχειρίδιο διακυβέρνησης. Παράλληλα, το 2020 εγκρίθηκε και το **Εγχειρίδιο Νομοπαρασκευαστικής Μεθοδολογίας**¹¹, το οποίο περιλαμβάνονται όλες οι αναγκαίες διαδικασίες για την τήρηση των αρχών που αφορούν στη νομοπαρασκευαστική διαδικασία και στην καλή νομοθέτηση. Ωστόσο, παρά τις προσπάθειες αυτές, οι οποίες υπήρξαν αποσπασματικές και ασυντόνιστες, η συζήτηση για την καλή νομοθέτηση στην Ελλάδα είναι πάντα υποβαθμισμένη και οι αρχές της καλής νομοθέτησης παραβιάζονται συστηματικά, καθώς η τυπολατρία, ο νομικισμός και ο ρυθμιστικός πληθωρισμός θεωρούνται, ακόμα και στις μέρες μας, δείγματα σωστής και επιτυχημένης διακυβέρνησης [2]. Επιπλέον, για να είναι εφικτή η εφαρμογή της καλής νομοθέτησης, θα πρέπει η αρχές της να εφαρμοστούν και στους ήδη υπάρχοντες νόμους, διαδικασία την οποία δεν έχει αναλάβει καμία κυβέρνηση λόγω του τεράστιου όγκου των νόμων αυτών.

Σύμφωνα με την **έρευνα**¹² του Κέντρου Φιλελεύθερων Μελετών (ΚΕΦΙΜ) για το έτος 2021, η ποιότητα της νομοθέτησης στην Ελλάδα παρουσίασε επιδείνωση 2,67 μονάδων στο Δείκτη Ποιότητας Νομοθέτησης σε σχέση με το 2020, γεγονός που οφείλεται «στις αρκετές τροποποιήσεις πρόσφατων νόμων, στην απουσία ποσοτικών δεδομένων στις εκθέσεις συνεπειών των νομοσχεδίων, και στον μικρό ποσοστό των αναγκαίων για την εφαρμογή των νόμων Υπουργικών Αποφάσεων που εκδίδονται εντός 6 μηνών από την ψήφιση του σχετικού νόμου». Στα ευρήματα της έρευνας περιλαμβάνονται τα εξής:

- Το 2021 ψηφίστηκαν πάνω από 11 χιλιάδες σελίδες νόμων. Μόνο το 55% των άρθρων που περιλήφθηκαν σε αυτούς τους νόμους είχε αναρτηθεί και ήταν ανοικτό στη διαβούλευση.
- Κατά το ίδιο έτος, μόνο 1 στους 5 νόμους προβλέπει απλούστευση διαδικασιών.
- Υπήρξε σημαντική αύξηση του συνόλου των τροπολογιών που κατατέθηκαν λιγότερες από 3 ημέρες πριν την ψήφιση τους.
- Αυξήθηκε σημαντικά ο συνολικός αριθμός των άρθρων όλων των τροπολογιών που ψηφίστηκαν κατά μέσο όρο ανά νόμο.
- Το ποσοστό των τροπολογιών που ψηφίζονται και είναι άσχετες με το κύριο αντικείμενο του νόμου που αφορούν έφτασε το 82,8%.
- Κατά την πενταετία 2017-2021, η ποιότητα της νομοθέτησης επιδεινώνεται σημαντικά στο τέλος κάθε έτους. Ειδικά τον μήνα Δεκέμβριο ψηφίζονται οι περισσότερες τροπολογίες, γεγονός που καταδεικνύει την προχειρότητα του νομοθετικού σχεδιασμού.

Στις αρχές του 2020, η διάδοση του SARS-COV-2 προκάλεσε την πανδημία του COVID-19 και έκανε επιτακτική και άμεση την ανάγκη για πλήρη, αποτελεσματική, λειτουργική και σύγχρονη ηλεκτρονική διακυβέρνηση. Έτσι, στα μέσα του 2020 έγινε διαθέσιμο στους πολίτες το σύστημα **gov.gr**¹³ που χρησιμοποιείται ακόμα και σήμερα και περιλαμβάνει μια ευρεία γκάμα υπηρεσιών με τελικό στόχο να αποτελέσει το κέντρο ψηφιακής εξυπηρέτησης που θα συγκεντρώνει όλες τις λύσεις και όσα πρέπει να γνωρίζουν οι πολίτες και οι επιχειρήσεις που συναλλάσσονται με το δημόσιο.

7 https://gslegal.gov.gr/?page_id=120

8 <https://www.kodiko.gr/nomothesia/document/64889/nomos-4048-2012>

9 <https://www.kodiko.gr/nomothesia/document/657744/p.d.-98-2020>

10 <https://www.kodiko.gr/nomothesia/document/545222/nomos-4622-2019>

11 <https://diavgeia.gov.gr/doc/60%CE%99%CE%9946%CE%9C%CE%93%CE%A87-%CE%A1%CE%95%CE%97?inline=true>

12 https://kefim.org/wp-content/uploads/2020/07/RQI-2021_Final.pdf

13 <https://www.gov.gr/info/about-us>

Το μεγαλύτερο μέρος της Ελληνικής Νομοθεσίας είναι διαθέσιμο σε όλους μέσω της επίσημης **ιστοσελίδας του Εθνικού Τυπογραφείου**¹⁴. Σε αυτή, ο χρήστης μπορεί να αναζητήσει συγκεκριμένα νομικά έγγραφα με βάση το ΦΕΚ τους, την ημερομηνία κυκλοφορίας τους ή το συνδυασμό έτους και τεύχους Εφημερίδας της Κυβέρνησης στην οποία δημοσιεύτηκαν. Υπάρχουν και άλλοι κρατικοί φορείς που παρέχουν την ελληνική νομοθεσία στο κοινό, όπως η **ιστοσελίδα της ΑΑΔΕ**¹⁵ (Ανεξάρτητη Αρχή Δημοσίων Εσόδων), η οποία περιλαμβάνει αρκετά φίλτρα αναζήτησης, και η **βιβλιοθήκη της ΑΑΔΕ**¹⁶, που περιέχει, εκτός από Νόμους, διαφόρων ειδών δημόσια έγγραφα. Σε όλες τις περιπτώσεις, οι νόμοι βρίσκονται ακόμα μόνο σε μορφή PDF και σε λίγες περιπτώσεις και σε html μορφή, με αποτέλεσμα η χρήση των ιστοσελίδων να περιορίζεται στην παρουσίαση των νόμων και η αναζήτηση συγκεκριμένων πληροφοριών να είναι ακόμα ευθύνη του πολίτη. Δε γίνεται καμία χρήση Διασυνδεδεμένων Δεδομένων, καθυστερώντας, έτσι, την ένταξη της νομοθεσίας μας στον Σημασιολογικό Ιστό.

Υπάρχουν, ωστόσο, και μη κρατικές πρωτοβουλίες για τη διάθεση της ελληνικής νομοθεσίας. Η ιστοσελίδα της Τράπεζας Νομικών Πληροφοριών **NOMOS**¹⁷ παρέχει με συνδρομή την ελληνική νομοθεσία, καθώς και άλλα νομικά έγγραφα, σε επαγγελματίες νομικούς (δικηγόρους, συμβολαιογράφους, δικαστές κ.λ.π.), σε συλλόγους νομικών αλλά και σε επίσημους φορείς (όπως το Γραφείο του Πρωθυπουργού, υπουργεία, ασφαλιστικά ταμεία, πανεπιστημιακά ιδρύματα, εταιρίες κ.λ.π.). Το **kodiko.gr**¹⁸ παρέχει σε εγγεγραμμένους χρήστες επί πληρωμή μέρος της ελληνικής νομοθεσίας σε html μορφή με δυνατότητα εύκολης πλοήγησης και συνεχώς εμπλουτίζεται. Το **Qualex**¹⁹ είναι άλλη μια ιστοσελίδα με εγγραφή επί πληρωμή, στην οποία έχουν δωρεάν πρόσβαση οι φοιτητές και τα υπόλοιπα ενεργά μέλη της Νομικής Σχολής του ΕΚΠΑ. Το Qualex παρέχει στους συνδρομητές του, πέρα από την ελληνική νομοθεσία, πλήθος άλλων σχετικών κειμένων, όπως εγχειρίδια, διοικητικά έγγραφα, δικόγραφα, συμβάσεις, άρθρα νομικών κ.λ.π.. Η επίσης συνδρομητική ιστοσελίδα **sakkoulasOnline**²⁰ εξειδικεύεται στο Ιδιωτικό, το Δημόσιο και το Ποινικό Δίκαιο. Τέλος, η ιστοσελίδα **ΙΣΟΚΡΑΤΗΣ**²¹ αποτελεί την Τράπεζα Νομικών Πληροφοριών του Δικηγορικού Συλλόγου Αθηνών και παρέχει δωρεάν πρόσβαση στην ελληνική νομοθεσία, ωστόσο τα νομικά έγγραφα βρίσκονται και σε αυτήν την περίπτωση σε PDF μορφή.

Παρατηρούμε πως ούτε ο δημόσιος αλλά ούτε και ο ιδιωτικός τομέας στη χώρα μας έχουν κινητοποιηθεί για τη δημιουργία κάποιας πλατφόρμας που χρησιμοποιεί τις τεχνολογίες του Σημασιολογικού Ιστού για τη διάθεση της ελληνικής νομοθεσίας στο ευρύ κοινό σε μορφή Διασυνδεδεμένων Δεδομένων, γεγονός που και θα διευκόλυνε την αναζήτησή της από τους πολίτες αλλά και θα έλυνε, εν μέρει, τα προβλήματα που προκαλεί η κακή ποιότητα νομοθέτησης που επικρατεί. Οι μόνες προσπάθειες που έχουν γίνει προς αυτή την κατεύθυνση αφορούν μικρής εμβέλειας πρωτοβουλίες, είτε ατομικές είτε μικρών ομάδων, οι οποίες, όπως είναι φυσικό, έχουν περιορισμένο πεδίο εφαρμογής.

1.4 Αντικείμενο και Στόχοι της Παρούσας Εργασίας

Όπως έχει περιγραφεί παραπάνω, παρότι έχουν γίνει προσπάθειες για την διάθεση της ελληνικής νομοθεσίας σε μορφή σημασιολογικών δεδομένων, αυτή τη στιγμή δεν υπάρχει ενεργή πλατφόρμα που να εξυπηρετεί την ανάγκη αυτή. Επιπλέον, με βάση τα Συμπεράσματα του Συμβουλίου της ΕΕ για την εισαγωγή του ELI (2012/C 325/02²²), από το 2012, η Ευρωπαϊκή Ένωση άρχισε να αναπτύσσει και, στη συνέχεια, να

14 <https://www.et.gr/>

15 <https://www.aade.gr/customs/elliniki-nomothesia>

16 <http://elib.aade.gr/elib>

17 <https://lawdb.intrasoftnet.com/portal/home>

18 <https://www.kodiko.gr/>

19 <https://www.qualex.gr/el-GR/>

20 <https://www.sakkoulas-online.gr/>

21 <https://www.dsanet.gr/>

22 [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52012XG1026\(01\)](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52012XG1026(01))

επεκτείνει την οντολογία ELI, έτσι ώστε όλα τα κράτη-μέλη της να δημοσιεύουν τα νομικά δεδομένα τους στον Σημασιολογικό Ιστό με κοινή μορφή, με στόχο τη διευκόλυνση της χρήσης και της ανταλλαγής της πληροφορίας μεταξύ τους.

Σκοπός της παρούσας εργασίας είναι η δημιουργία οντολογίας για την ελληνική νομοθεσία μέσω της επέκτασης του ευρωπαϊκού προτύπου ELI έτσι ώστε να εξυπηρετήσει τις ιδιαιτερότητές της. Επιμέρους στόχοι της εργασίας αποτελούν:

- Η μελέτη, κατανόηση και εξήγηση των βασικών εννοιών και τεχνολογιών του Σημασιολογικού Ιστού.
- Η μελέτη πάνω στη χρήση των Διασυνδεδεμένων Δεδομένων στο νομικό πεδίο παγκοσμίως.
- Ο εντοπισμός των ιδιαιτεροτήτων του ελληνικού νομικού συστήματος.
- Η προσαρμογή και επέκταση του ευρωπαϊκού αναγνωριστικού ELI με στόχο την ανάπτυξη οντολογίας κατάλληλης για τα ελληνικά νομικά δεδομένα.
- Η μοντελοποίηση νομικών κειμένων σε μορφή RDF.
- Η παρουσίαση σεναρίων χρήσης της οντολογίας που θα αναπτυχθεί με τη χρήση SPARQL ερωτημάτων πάνω στα σημασιολογικά έγγραφα RDF.

1.5 Σχετικές Εργασίες

Τα τελευταία χρόνια έχουν γίνει πολλές εργασίες που αφορούν τη δημιουργία οντολογιών και την ανάπτυξη πλατφορμών για τη δημοσίευση νομικών δεδομένων στον Σημασιολογικό Ιστό, τόσο παγκοσμίως όσο και στην Ελλάδα. Παρακάτω παρουσιάζουμε μερικές από αυτές:

Διεθνείς Εργασίες

Το 2014, οι Hyvönen Eero, et al. [3] ανέπτυξαν ένα μοντέλο και την αντίστοιχη πλατφόρμα (**LDF.fi**²³) για τη δημοσίευση της φινλανδικής νομοθεσίας σε μορφή Διασυνδεδεμένων Δεδομένων. Η πλατφόρμα αυτή, πέρα από τις δυνατότητες του Σημασιολογικού Ιστού, προσφέρει δύο επιπλέον πλεονεκτήματα στο χρήστη: Την επεξήγηση του λεξιλογίου που χρησιμοποιεί κάθε σύνολο δεδομένων που δημοσιεύεται σε αυτήν και την αξιολόγηση της ποιότητας των δεδομένων αυτών. Η επεξήγηση του λεξιλογίου γίνεται αυτόματα μέσω του διαδικτυακού εργαλείου επεξήγησης λεξιλογίων **LODE**²⁴, ενώ η αξιολόγηση των δεδομένων επιτυγχανόταν μέσω της (ανενεργούς πλέον) υπηρεσίας vocab.at, η οποία ανέλυε το σύνολο των δεδομένων και παρήγαγε μια αναφορά σε HTML μορφή, όπου επισημαίνονταν τυχόν λάθη και ανακρίβειες.

Σε συνέχεια της εργασίας αυτής, οι ίδιοι συγγραφείς δημιούργησαν το 2020 [4] μια νέα πλατφόρμα, το **SEMANTIC FINLEX**²⁵, βασισμένη στην LDF.fi. Η νέα αυτή πλατφόρμα ακολουθεί το πρότυπο ELI. Επιπλέον, άρχισαν από το 2019 [5] να αναπτύσσουν και μια σημασιολογική πύλη ονομαζόμενη LawSampo. Η πύλη αυτή προσφέρει στο χρήστη τη δυνατότητα να χρησιμοποιήσει τα διασυνδεδεμένα δεδομένα του SEMANTIC FINLEX σε διάφορα σενάρια χρήσης και είναι βασισμένη βασισμένη στο μοντέλο Sampo [6].

Το 2018, οι Victor Rodríguez-Doncel et al. [7] μετέτρεψαν την ισπανική νομοθεσία σε RDF σημασιολογικά έγγραφα. Για να το πετύχουν αυτό, ανέπτυξαν ένα μοντέλο δεδομένων προσαρμοσμένο στα δεδομένα της

23 <https://www.ldf.fi/>

24 <https://essepuntato.it/lode/>

25 <https://data.finlex.fi/fi/main>

Ισπανίας και βασισμένο στο πρότυπο ELI. Με τη βοήθεια δικών τους script, κατέβασαν τμήματα της ισπανικής νομοθεσίας από την επίσημη [ιστοσελίδα](#)²⁶ που αυτή δημοσιεύεται σε μορφή XML αρχείων και τα μετέτρεψαν σε αρχεία με RDF μορφή με βάση το προαναφερθέν μοντέλο δεδομένων. Στη συνέχεια, συνδέσαν τα δεδομένα αυτά με σημασιολογικά δεδομένα από άλλες RDF βάσεις δεδομένων.

Το 2021 πραγματοποιήθηκε η εργασία [8], στην οποία οι συγγραφείς περιέγραψαν μια μεθοδολογία μετατροπής της αυστριακής νομοθεσίας σε Διασυνδεδεμένα Δεδομένα, με στόχο την ένταξή της σε έναν πανευρωπαϊκό γράφο γνώσης νομικών πληροφοριών βασισμένο στο πρότυπο ELI. Δημιούργησαν το AustroVoc, ένα λεξιλόγιο που επεκτείνει το πρότυπο ELI με τις κατάλληλες αυστριακές ορολογίες και, στη συνέχεια, τον εμπλούτισαν με δεδομένα προερχόμενα και από τα δομημένα δεδομένα της ιστοσελίδας RIS, αλλά και από τα μη δομημένα δεδομένα της, τα οποία επεξεργάστηκαν με Εργαλεία Ανάλυσης Φυσικής Γλώσσας (Natural Language Processing Tools - NLP Tools) έτσι ώστε να εξάγουν τα κατάλληλα μεταδεδομένα.

Ελληνικές Εργασίες

Το 2023, στην εργασία [9] παρουσιάστηκε μια αντιστοίχιση των οντολογιών Akoma Ntoso και ELI. Η αντιστοίχιση έγινε από τους ερευνητές που εργάστηκαν στην εργασία αυτή και αξιολογήθηκε από ειδικούς στον τομέα της παρουσίασης νομικών πληροφοριών. Πέρα από τη μελέτη της συμβατότητας των δύο αυτών οντολογιών, η εργασία αυτή προτείνει μια μεθοδολογία για την περιγραφή και την αντιστοίχιση νομικών οντολογιών, με τελικό στόχο τη διαλειτουργικότητά τους.

Όσον αφορά τις προσπάθειες για το ελληνικό σύστημα δημοσίευσης νομικών δεδομένων, το 2013 [10] παρουσιάστηκε μια σειρά βημάτων από την οποία θα μπορούσαν να εξαχθούν διασυνδεδεμένα δεδομένα για την ελληνική νομοθεσία από την κυβερνητική ψηφιακή πύλη ΕΡΜΗΣ. Παρότι η πύλη αυτή δε βρίσκεται πλέον σε λειτουργία, τα προβλήματα που περιγράφονται θυμίζουν αρκετά τη σημερινή πραγματικότητα και τα βήματα που προτείνονται (Ανάκτηση Δεδομένων - Χρήση Οντολογιών για τη Μοντελοποίησή τους σε Διασυνδεδεμένα Δεδομένα - Συνδέσεις με άλλα Σύνολα Δεδομένων - Στατιστική Ανάλυση Αποτελεσμάτων Χρήσης) για τη δημοσίευση των ελληνικών νομικών πληροφοριών στο Σημασιολογικό Ιστό αποτελούν χρήσιμο οδηγό για μελλοντικές προσπάθειες πάνω σε αυτόν τον τομέα.

Το 2016, οι Koniaris Marios et al [11] προέτειναν μια μέθοδο απόκτησης σημασιολογικών μεταδεδομένων από ελληνικά νομικά αρχεία μη δομημένης μορφής. Αφού προσδιόρισαν τη βασική μορφή που έχουν τα περισσότερα ελληνικά νομικά έγγραφα, χρησιμοποίησαν στα αντίστοιχα αρχεία κατάλληλους εξαγωγείς (extractors) κειμένου και εικόνων και, στη συνέχεια, αναλυτές κειμένου (parsers) στα εξαγόμενα δεδομένα έτσι ώστε να εντοπίσουν και να αντλήσουν τις απαραίτητες πληροφορίες. Στη συνέχεια, κατηγοριοποίησαν τις πληροφορίες αυτές και τις μοντελοποίησαν σε σημασιολογικά έγγραφα που ακολουθούν το πρότυπο Akoma Ntoso (που πλέον ονομάζεται LegalDocML). Για τη δημοσίευση των σημασιολογικών αυτών εγγράφων περιέγραψαν την πλατφόρμα Σόλων [12], η οποία μοντελοποιεί, με βάση τη προηγούμενη μέθοδο, νομικά έγγραφα τα οποία εισάγονται σε αυτή είτε χειροκίνητα είτε αυτόματα από συγκεκριμένες πηγές. Η πλατφόρμα αυτή, επιπλέον, παρέχει στους χρήστες ποικίλες δυνατότητες για την αναζήτηση νομικών κειμένων. Ο μηχανισμός αυτός εφαρμόστηκε στην πλατφόρμα της ΑΑΔΕ, [e-Lib](#)²⁷, με στόχο την παροχή σημασιολογικών δεδομένων για τα νομικά έγγραφα που περιέχονται σε αυτή, ωστόσο δεν έχει χρησιμοποιηθεί στα μεταγενέστερα νομικά της αρχεία.

Επιπλέον, το 2017 οι Chalkidis Ilias, et al. [13] ανέπτυξαν την OWL οντολογία «Nomothesia» με στόχο τη μοντελοποίηση του περιεχομένου της ελληνικής νομοθεσίας. Χρησιμοποιώντας το Greek Government Gazette Parser, ένα εργαλείο ανάλυσης κειμένου δικής τους κατασκευής, μετέτρεψαν όλα τα τεύχη της Εφημερίδας της Κυβέρνησης από το έτος 2006 ως το έτος 2015 σε Διασυνδεδεμένα Δεδομένα που

²⁶ <https://www.boe.es/>

²⁷ <http://www.publicrevenue.gr/elib/>

ακολουθούν το πρότυπο RDF και τα δημοσίευσαν στην [πλατφόρμα Nomothesi@](#)²⁸ (η οποία σταμάτησε να λειτουργεί κατά τη διάρκεια διεκπεραίωσης της παρούσας εργασίας) που επίσης κατασκεύασαν [14]. Η πλατφόρμα αυτή χρησιμοποιεί RESTful API για τη διάθεση των νομικών κειμένων που βρίσκονται σε αυτή σε ποικίλες μορφές (JSON, XML κ.λ.π.) και προσφέρει στους χρήστες της τη δυνατότητα να κάνουν σύνθετες αναζητήσεις στο σύνολο των δεδομένων που φιλοξενεί, ενώ διαθέτει SPARQL endpoint στην περίπτωση που ο χρήστης θέλει να θέσει κάποιο SPARQL ερώτημα στα δεδομένα αυτά. Το 2018, σε μεταγενέστερη εργασία [15] τους, μέσω της χρήσης νευρωνικών δικτύων με σκοπό την ονομαστική αναγνώριση οντοτήτων, εμπλούτισαν το γράφο γνώσης της ελληνικής νομοθεσίας με περισσότερα δεδομένα που αφορούν πρόσωπα, οργανισμούς, γεωπολιτικές και γεωχωρικές οντότητες και αναφορές σε νομοθεσία ή σε δημόσια κείμενα. Για την μοντελοποίηση των νέων αυτών δεδομένων εισήγαγαν ένα νέο RDF λεξιλόγιο και, όπου ήταν δυνατό, συνέδεσαν τα δεδομένα αυτά με ήδη υπάρχοντα Διασυνδεδεμένα Δεδομένα.

1.6 Οργάνωση Κειμένου

Η παρούσα Διπλωματική Εργασία έχει την εξής δομή:

Στο Κεφάλαιο 2 αναλύουμε τις βασικές έννοιες και τεχνολογίες που αφορούν το Σημασιολογικό Ιστό, κάνοντας ιδιαίτερη αναφορά στο πρότυπο ELI. Στο Κεφάλαιο 3 παρουσιάζουμε το ELIgr, δηλαδή την επέκταση του προτύπου ELI για τα ελληνικά δεδομένα, μαζί με ένα παράδειγμα σημασιολογικού εγγράφου δημιουργημένο με βάση την επέκταση αυτή που ακολουθεί το πρότυπο RDF. Στο Κεφάλαιο 4 αναλύουμε τη διαδικασία αυτόματης και μαζικής μετατροπής των LegalDocML αρχείων σε RDF σημασιολογικά έγγραφα. Στο Κεφάλαιο 5 παρουσιάζουμε παραδείγματα χρήσης του μοντέλου μας μέσω SPARQL ερωτημάτων. Τέλος, στο Κεφάλαιο 6 συνοψίζουμε την εργασία μας, αναλύουμε συμπεράσματά που προέκυψαν από αυτή και περιγράφουμε μελλοντικές επεκτάσεις της.

28 <https://legislation.di.uoa.gr/>

2

Θεωρητικό Υπόβαθρο

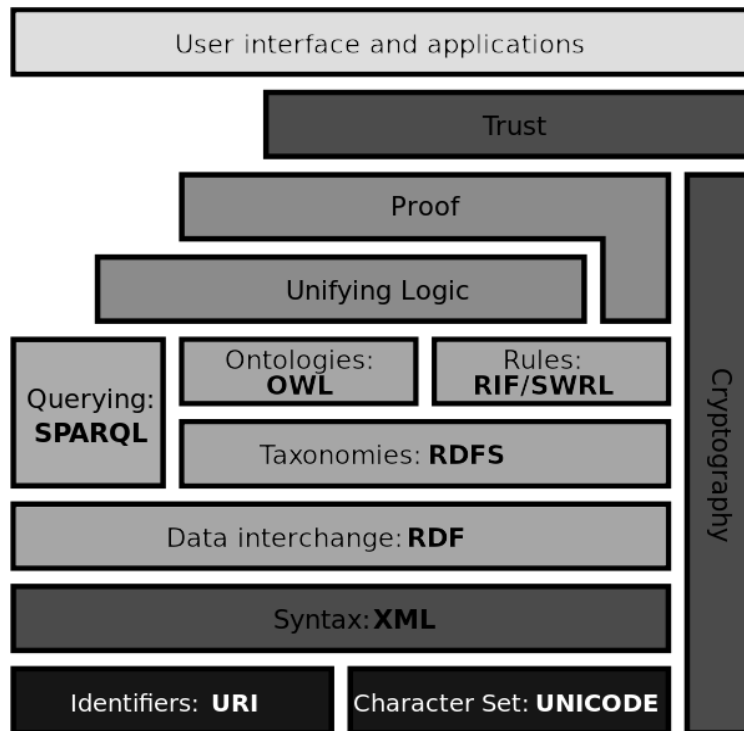
2.1 Σημασιολογικός Ιστός (Semantic Web) και Διασυνδεδεμένα Δεδομένα (Linked Data)

Για την καλύτερη κατανόηση της παρούσας διπλωματικής εργασίας κρίνεται απαραίτητο να γίνει μια εισαγωγή στις βασικές έννοιες που αφορούν τον Σημασιολογικό Ιστό και τα Διασυνδεδεμένα Δεδομένα.

2.1.1 Σημασιολογικός Ιστός (Semantic Web)

Ο Σημασιολογικός Ιστός [16], γνωστός και ως Web 3.0, αποτελεί μια επέκταση του Παγκοσμίου Ιστού η οποία μοντελοποιεί τα δεδομένα δίνοντάς τους δομή και μετατρέποντας την πληροφορία σε γνώση. Ονομάζεται έτσι διότι δίνει έμφαση στη σημασία της δημοσιευμένης πληροφορίας. Αυτό επιτυγχάνεται με τον ορισμό μεταδεδομένων (metadata) τα οποία θα περιέχονται στην πληροφορία, θα είναι κοινά για όλους και θα «κατανοούνται» από τις μηχανές, με αποτέλεσμα την ευκολότερη και αποδοτικότερη αναζήτηση, επεξεργασία και χρήση της πληροφορίας.

Η αρχιτεκτονική του Σημασιολογικού Ιστού είναι ουσιαστικά μια ενίσχυση της αρχιτεκτονικής του σημερινού Παγκοσμίου Ιστού με περισσότερα επίπεδα. Τα επίπεδα αυτά αποτελούν σύνολα τεχνολογιών και γλωσσών προγραμματισμού που βασίζονται, με τη σειρά τους, σε κατώτερα επίπεδα. Μια απεικόνιση της αρχιτεκτονικής του Σημασιολογικού Ιστού από τον εμπνευστή του, Tim Berners-Lee, γνωστή και ως «Στοιβά του Σημασιολογικού Ιστού» είναι η παρακάτω:



Εικόνα 1: Στοιβά του Σημασιολογικού Ιστού

Τα επίπεδα που απεικονίζονται, από κάτω προς τα πάνω, είναι τα ακόλουθα:

Επίπεδο 1 (URI, UNICODE): Αποτελούν την ήδη υπάρχουσα δομή του παγκοσμίου ιστού και είναι τα πρότυπα με βάση τα οποία αναπαρίστανται οι χαρακτήρες και ταυτοποιούνται οι πηγές πληροφορίας.

Επίπεδο 2 (XML): Με βάση το πρότυπο XML, μιας μεταγλώσσας σήμανσης που δίνει τη δυνατότητα δημιουργίας πληροφοριακών πόρων με πολύπλοκη και ευέλικτη δομή, επιτυγχάνεται η αναπαράσταση των δεδομένων, χωρίς όμως η αναπαράσταση αυτή να συνδέεται με τη σημασία τους.

Επίπεδο 3 (RDF): Εδώ γίνεται η αναπαράσταση νοήματος και η περιγραφή των μεταδεδομένων και των σχέσεων μεταξύ των πηγών πληροφορίας, ώστε να είναι κατανοητά από τους ηλεκτρονικούς υπολογιστές.

Επίπεδο 4 (RDFs): Η RDFs είναι μια γλώσσα προγραμματισμού που περιγράφει κλάσεις και σχέσεις μεταξύ πηγών πληροφορίας με βάση το πρότυπο RDF. Παρέχει το κατάλληλο σημασιολογικό πλαίσιο για την εξαγωγή συμπερασμάτων.

Επίπεδο 5 (OWL, RIF/SWRL, SPARQL): Οι γλώσσες OWL, RIF και SWRL μας προσφέρουν τη δυνατότητα να θέσουμε κανόνες και περιορισμούς στην αναπαράσταση πληροφοριών, δίνοντάς μας περισσότερη ευελιξία. Με τη γλώσσα SPARQL θέτουμε ερωτήματα στα σημασιολογικά δεδομένα είτε για να ελέγξουμε την ορθότητά τους, είτε για να πάρουμε τις πληροφορίες που θέλουμε..

Επίπεδα 6 και 7 (Unifying Logic and Proof): Εδώ δίνεται η δυνατότητα αναπαράστασης κανόνων κατηγορηματικής λογικής και συσχετίσεων μεταξύ των δεδομένων μιας οντολογίας για την εξαγωγή συμπερασμάτων έπειτα από λογική ανάλυση μέσω του υπολογιστή. Χάρη στους κανόνες αυτούς ο υπολογιστής μπορεί να ελέγξει αν τα δεδομένα που προσθέτουμε τηρούν τους κανόνες και τις απαιτήσεις της εκάστοτε οντολογίας.

Επίπεδο 8 (Trust): Ακόμα και σήμερα δεν έχει πλήρως αναπτυχθεί και αφορά στην υλοποίηση μηχανισμών ικανών να κατηγοριοποιούν το βαθμό αξιοπιστίας και να αξιολογούν τις πληροφορίες που παρέχουν οι διαδικτυακές πηγές βάσει ψηφιακών υπογραφών και πιστοποιητικών.

Η Κοινοπραξία του Παγκοσμίου Ιστού (W3C) είναι μια οργάνωση της οποίας οι οργανώσεις-μέλη συνεργάζονται με στόχο την ανάπτυξη των προτύπων του Παγκοσμίου Ιστού. Ασχολείται επίσης με την ανάπτυξη λογισμικού για τον Σημασιολογικό Ιστό και έχει αναλάβει την τυποποίησή του. Η τυποποίηση αυτή περιλαμβάνει ένα σύνολο κανόνων και τεχνολογιών που παρέχουν τη δυνατότητα περιγραφής εννοιών,

όρων και σχέσεων μέσα σε πεδία γνώσης. Στις τεχνολογίες αυτές, που ορίζονται ως W3C Τυποποίηση, περιλαμβάνονται τα RDF (γενική μέθοδος περιγραφής πληροφοριών), RDFS, SKOS, SPARQL (RDF γλώσσα ερωτημάτων), N3, N-Triples (μορφή για αποθήκευση και μετάδοση πληροφοριών), Turtle, OWL (οικογένεια γλωσσών αναπαράστασης γνώσης), RIF, JSON-LD και ActivityPub (απλός τρόπος επικοινωνίας client και server). Μερικές από αυτές τις τεχνολογίες θα αναλυθούν εκτενέστερα στην Ενότητα 2.2.

2.1.2 Διασυνδεδεμένα Δεδομένα (Linked Data)

Στον Παγκόσμιο Ιστό, τα δεδομένα είναι αποθηκευμένα σε ετερογενείς βάσεις δεδομένων, με αποτέλεσμα η σύνδεση και η ανταλλαγή τους με στόχο εξαγωγή συμπερασμάτων να καθίσταται συχνά δύσκολη. Ο όρος Διασυνδεδεμένα Δεδομένα [17] περιγράφει μία μέθοδο δημοσιοποίησης δομημένων δεδομένων ώστε να είναι αλληλένδετα και να γίνουν πιο χρήσιμα. Η μέθοδος αυτή στηρίζεται στις γνωστές τεχνολογίες του Ιστού, όπως το HTTP και τα URIs, αλλά αντί να τις χρησιμοποιεί μόνο για να εξυπηρετεί ιστοσελίδες για τους ανθρώπινους αναγνώστες, τις επεκτείνει ώστε να ανταλλάσσουν πληροφορίες με τρόπο που θα μπορούν να διαβαστούν αυτόματα από τους υπολογιστές. Αυτό επιτρέπει δεδομένα από διαφορετικές πηγές να συνδέονται και να μπορούν να αναζητηθούν. Στον Παγκόσμιο Ιστό, αυτό που συνηθίζεται είναι να συνδέονται αρχεία δεδομένων μεταξύ τους με υπερσυνδέσμους (hypertext) μέσω της γλώσσας HTML (HyperText Markup Language). Τα Διασυνδεδεμένα Δεδομένα χρησιμοποιούν τη δομή RDF ώστε να συνδέσουν πιο «αυθαίρετες» έννοιες με αποτέλεσμα τη δημιουργία ενός «Ιστού Δεδομένων» (Web of Data), ο οποίος περιγράφεται σαν ένας «ιστός των πραγμάτων στον κόσμο, περιγραφόμενος από δεδομένα στο διαδίκτυο». Όσο περισσότερα δεδομένα έχει ο ιστός αυτός, τόσο πιο ισχυρός είναι.

Ο Τιμ Μπέρνερς-Λι όρισε, το 2006, τις 4 αρχές που πρέπει να διέπουν τα Διασυνδεδεμένα Δεδομένα:

1. Χρησιμοποιούμε τα URIs (Uniform Resource Identifier) για να προσδιορίσουμε πράγματα.
2. Χρησιμοποιούμε τα HTTP (Hypertext Transfer Protocol) URIs ώστε αυτά τα πράγματα να μπορούν να αναφερθούν και να αναζητηθούν από ανθρώπους και από πράκτορες χρηστών.
3. Δίνουμε χρήσιμες πληροφορίες σχετικά με το θέμα όταν το URI του ορίζεται εκ νέου χρησιμοποιώντας τυποποιημένους τύπους όπως RDF/XML/SPARQL.
4. Συμπεριλαμβανούμε συνδέσεις σε άλλα σχετικά URIs με τα εκτεθειμένα δεδομένα ώστε να βελτιωθεί η ανακάλυψή τους από άλλες σχετικές πληροφορίες στον Ιστό και ώστε όσοι τα αναζητούν να ανακαλύψουν περισσότερα δεδομένα.

Στο **TED συνέδριο του 2009**²⁹, ο Τιμ Μπέρνερς-Λι παρουσίασε 3 βασικούς κανόνες για τα Διασυνδεδεμένα Δεδομένα:

1. Όλα τα είδη των εννοιολογικών πραγμάτων έχουν ονόματα που ξεκινούν με HTTP.
2. Όταν ψάχνω ένα όνομα που ξεκινά με HTTP, παίρνω πίσω κάποια δεδομένα σε μια τυποποιημένη μορφή (χωρίς να είναι απαραίτητο η μορφή αυτή να είναι RDF ή XML). Τα δεδομένα αυτά είναι χρήσιμες πληροφορίες τις οποίες κάποιος θα ήθελε να ξέρει σχετικά με αυτό το οποίο αναζητεί.
3. Για κάθε τι που αναζητώ με ένα HTTP όνομα υπάρχουν κάποιες σχέσεις. Κάθε φορά που το αντικείμενο αναζήτησής μου εκφράζει μια σχέση, τότε το άλλο πράγμα με το οποίο είναι συνδεδεμένο με τη σχέση αυτή, έχει και αυτό ένα όνομα που ξεκινά με HTTP, ώστε να μπορώ, αντίστοιχα, να βρω πληροφορίες και για αυτό.

Τα Διασυνδεδεμένα Δεδομένα είναι περισσότερο «αποδοτικά» όταν είναι ανοιχτά. Με τον όρο Ανοιχτά Δεδομένα εννοούμε δεδομένα ελεύθερα διαθέσιμα για χρήση, επεξεργασία, διαμοιρασμό και εκμετάλλευση για οποιονδήποτε σκοπό χωρίς περιορισμούς. Αξιολογούνται, κατά τον Τιμ Μπέρνερς-Λι, ανάλογα με την ποιότητά τους σε 5 κατηγορίες, με την υψηλότερη από αυτές να είναι τα Ανοιχτά Διασυνδεδεμένα Δεδομένα. Ο Τιμ Μπέρνερς-Λι όρισε τα Ανοιχτά Διασυνδεδεμένα Δεδομένα (Linked Open Data) ως Διασυνδεδεμένα Δεδομένα που είναι προσβάσιμα σε όλους με ανοιχτή άδεια, ώστε να μην εμποδίζεται η δωρεάν επαναχρησιμοποίησή τους. Γνωστά σύνολα Ανοιχτών Διασυνδεδεμένων Δεδομένων αποτελούν τα

29 https://web.archive.org/web/20110607085246/http://video.ted.com/talks/podcast/TimBerners-Lee_2009.mp4

Dbpedia³⁰, **Wikibase**³¹ και **Wikidata**³².

Αξίζει, επίσης, αναφοράς, το Linking Data Open Project. Το Linking Data Open Project είναι μια πρωτοβουλία που ξεκίνησε από τον Ιανουάριο του 2007 με τη στήριξη του W3C και έχει σκοπό το σταδιακό «χτίσιμο» του Ιστού Δεδομένων. Αυτό θα επιτευχθεί με την εύρεση συλλογών δεδομένων (datasets) που έχουν δημοσιευτεί με ανοιχτή άδεια, την μετατροπή τους σε μορφή RDF σύμφωνα με τις αρχές των Διασυνδεδεμένων Δεδομένων και την δημοσίευσή τους στο διαδίκτυο. Η εργασία αυτή, που ξεκίνησε από ερευνητές σε πανεπιστήμια και μικρές εταιρίες, έχει πλέον τη στήριξη μεγάλων οργανισμών όπως ο BBC και η Βιβλιοθήκη του Κογκρέσου των Η.Π.Α.. Αυτό επιτεύχθηκε διότι καθένας μπορεί να συνεισφέρει στην εργασία αυτή, εφόσον δημοσιεύει συλλογές δεδομένων σύμφωνα με τις αρχές των Διασυνδεδεμένων Δεδομένων, και εφόσον τις συνδέει με ήδη υπάρχουσες συλλογές δεδομένων.

Αν και είναι δύσκολο να μετρηθεί με ακρίβεια, υπολογίζεται πως ο Ιστός Δεδομένων του Linked Data Open Project διέθετε, τον Οκτώβριο του 2007, πάνω από 2 δισεκατομμύρια RDF τριάδες συνδεδεμένες με πάνω από 2 εκατομμύρια RDF σχέσεις. Το μέγεθος αυτό αυξήθηκε, έως τον Σεπτέμβριο του 2011 σε πάνω από 31 δισεκατομμύρια RDF τριάδες με περίπου 504 εκατομμύρια σχέσεις. Σύμφωνα με **αναφορά του 2014**³³, περίπου το 51% των συλλογών δεδομένων στο Linked Data Open Project αφορά δεδομένα από κοινωνικά δίκτυα, ενώ αμέσως επόμενα έρχονται τα κυβερνητικά δεδομένα, με ποσοστό περίπου 18%.

Παρατηρούμε ότι, ενώ έχουμε 15πλασιασμό των RDF τριάδων, οι μεταξύ τους συνδέσεις έχουν, στο ίδιο διάστημα, 250πλασιαστεί. Αυτό συμβαίνει διότι οι νέες συλλογές δεδομένων που προστίθενται δεν έχουν μόνο έναν δεδομένο αριθμό σχέσεων μεταξύ των δικών τους RDF τριάδων, αλλά έχουν πολλές σχέσεις και με RDF τριάδες από τις ήδη υπάρχουσες συλλογές δεδομένων. Επομένως, εφόσον όσο προστίθενται νέα δεδομένα στον Ιστό Δεδομένων δημιουργούνται περισσότερες σχέσεις με αυτά και τα ήδη υπάρχοντα δεδομένα, είναι λογικό ο ρυθμός αύξησης των μεταξύ τους σχέσεων να είναι πολύ μεγαλύτερος από τον ρυθμό αύξησης των ίδιων των δεδομένων.

Οι παρακάτω εικόνες αποτελούν τους γράφους γνώσης (knowledge graphs) του Ιστού Δεδομένων του Linked Data Open Project (οι γράφοι αυτοί ονομάζονται Linked Open Data Cloud (LOD Cloud)) για το έτος 2014 και το έτος 2022 αντίστοιχα. Βλέπουμε, ξανά, την τεράστια αύξηση του όγκου των δεδομένων σε κάθε διαφορετικό πεδίο γνώσης (κάθε πεδίο γνώσης αναπαρίσταται με διαφορετικό χρώμα), αλλά και την ακόμα μεγαλύτερη αύξηση των μεταξύ τους σχέσεων τα τελευταία χρόνια, σε βαθμό που στην δεύτερη εικόνα οι συνδέσεις είναι τόσες πολλές που οι γραμμές που τις αναπαριστούν φαίνονται σαν μια μεγάλη μαύρη σκιά και απαιτείται τεράστια μεγέθυνση (η οποία δεν θα επέτρεπε την πλήρη προβολή της εικόνας) για να τις ξεχωρίσουμε. Οι γράφοι αυτοί και τα Διασυνδεδεμένα Δεδομένα που περιλαμβάνουν είναι διαθέσιμα στην **ιστοσελίδα του Linked Open Data Cloud**³⁴.

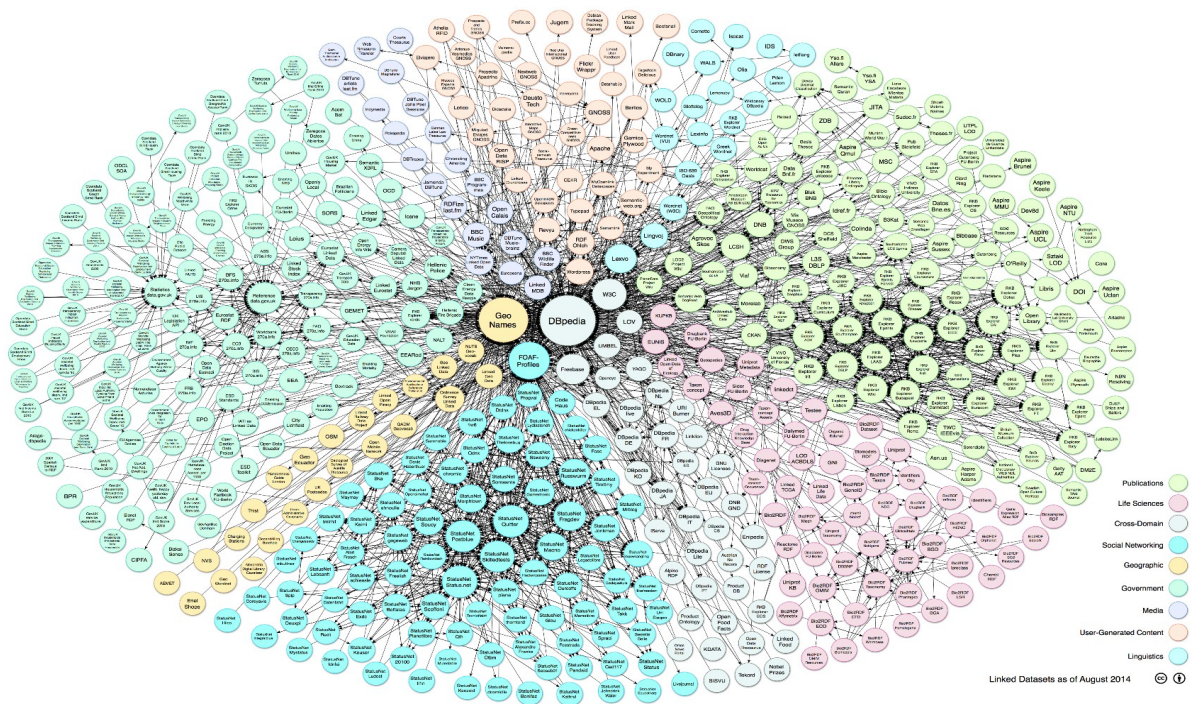
30 <https://www.dbpedia.org/>

31 <https://wikiba.se/>

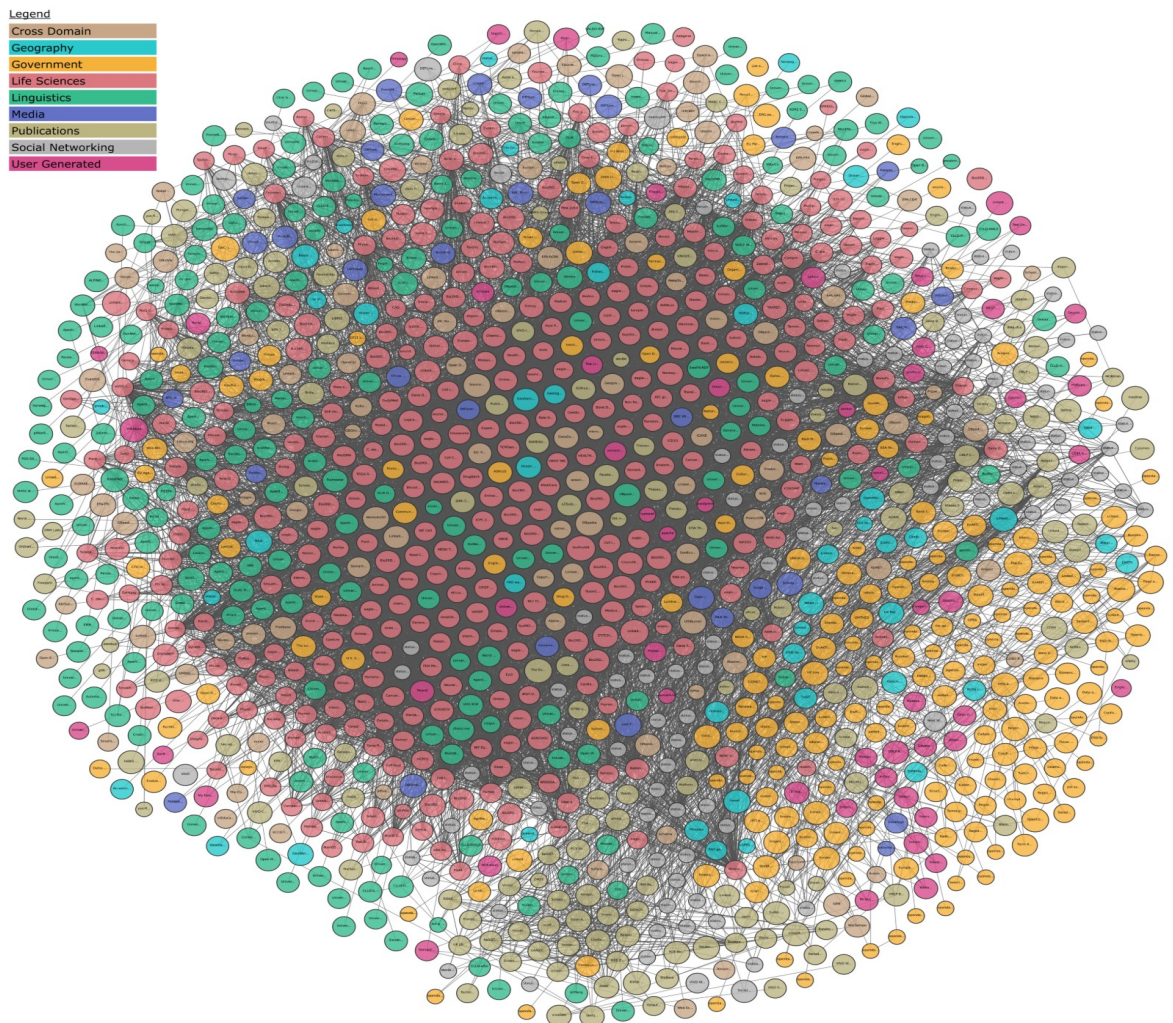
32 https://www.wikidata.org/wiki/Wikidata:Main_Page

33 <http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/>

34 <http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/>



Εικόνα 2: Το LOD Cloud του έτους 2014



Εικόνα 3: Το LOD Cloud του έτους 2022

2.1.3 Οντολογίες (Ontologies)

Για να μπορεί να χρησιμοποιηθεί ο Σημασιολογικός Ιστός, είναι απαραίτητο να υπάρχουν δεδομένα μοντελοποιημένα σε οντολογίες [18]. Υπάρχουν τριών τύπων οντολογίες: οι οντολογίες πεδίου (domain ή top-level ontologies), οι ανώτερες οντολογίες (upper ontologies) και οι υβριδικές οντολογίες (hybrid ontologies), οι οποίες είναι ένας συνδυασμός των προηγούμενων δύο τύπων.

Η οντολογία πεδίου ή απλώς οντολογία στην πληροφορική είναι ένας τυπικός και σαφής ορισμός μιας κοινής και συμφωνημένης εννοιολογικής μορφοποίησης που αφορά ένα πεδίο ενδιαφέροντος. Αυτή η τυπική αναπαράσταση γνώσης ως ένα σύνολο εννοιών, σχέσεων και ιδιοτήτων μπορεί να χρησιμοποιηθεί για εξαγωγή συμπερασμάτων και νέας γνώσης και για την δομημένη περιγραφή γνώσης ενός πεδίου ενδιαφέροντος. Υπάρχουν, για παράδειγμα, οντολογίες για το φαγητό, για την ιατρική, για το νομικό πλαίσιο κ.λ.π..

Οι ανώτερες οντολογίες (π.χ. Dublin Core) είναι μοντέλα που ορίζουν σχέσεις και αντικείμενα που συναντώνται συχνά σε πολλές διαφορετικές οντολογίες πεδίου. Συνήθως χρησιμοποιούν λεξιλόγια (vocabularies) που περιλαμβάνουν έννοιες και περιγραφές αντικειμένων όπως αυτά ορίζονται στις διαφορετικές οντολογίες. Αποτελούν, ουσιαστικά, τον «συνδυαστικό κρίκο» μεταξύ διαφορετικών οντολογιών πεδίου. Επειδή οι διαφορετικές οντολογίες πεδίου αναπτύσσονται από διαφορετικά άτομα ή ομάδες, είναι συχνό το φαινόμενο να συναντούμε οντολογίες που αφορούν το ίδιο πεδίο και, παρόλα αυτά, είναι ασύμβατες μεταξύ τους. Αυτό συμβαίνει κυρίως διότι κάθε σχεδιαστής οντολογίας έχει τη δική του γλώσσα και κουλτούρα, με αποτέλεσμα να ορίζει, να κατανοεί και άρα να μοντελοποιεί διαφορετικά κάποιες έννοιες. Επιπλέον, διαφορές προκύπτουν και από τους διαφορετικούς σκοπούς χρήσης κάθε οντολογίας. Η κατάσταση αυτή οδηγεί όμως σε προβλήματα όταν πρέπει να συνδυαστούν οι οντολογίες για κάποιο κοινό έργο. Αν οι διαφορετικές οντολογίες δεν έχουν προκύψει από κάποια ανώτερη, η συγχώνευσή τους συνήθως είναι ακριβή και χρονοβόρα, καθώς πρέπει να γίνει, κατά κύριο λόγο, «χειροκίνητα». Αντιθέτως, αν οι διαφορετικές οντολογίες έχουν τις «ρίζες» τους σε κάποια ανώτερη, η συγχώνευσή τους αυτοματοποιείται και διευκολύνεται σε μεγάλο βαθμό.

Οι οντολογίες και τα δεδομένα είναι αλληλένδετα μεταξύ τους. Χωρίς τις οντολογίες, ο υπολογιστής δεν είναι σε θέση να γνωρίζει τις σχέσεις μεταξύ των δεδομένων που υπάρχουν άτακτα και ανοργάνωτα στον Παγκόσμιο Ιστό, καθώς και το ποιά δεδομένα πρέπει να αλληλεπιδράσουν με ποιά άλλα. Αντίστοιχα, μια οντολογία χωρίς δεδομένα δεν έχει κανένα νόημα, διότι, όντας μόνο μια εννοιολογική μορφοποίηση, ένα «καλούπι» χωρίς περιεχόμενο, δε μπορεί προσφέρει πληροφορία και γνώση. Ο συνδυασμός δεδομένων και οντολογιών είναι αυτός που δημιουργεί τον Σημασιολογικό Ιστό, ένα οργανωμένο παγκόσμιο σύστημα χάρη στο οποίο μπορούμε να φτάσουμε από τη μία πληροφορία στην άλλη ακολουθώντας τις σχέσεις που τις συνδέουν.

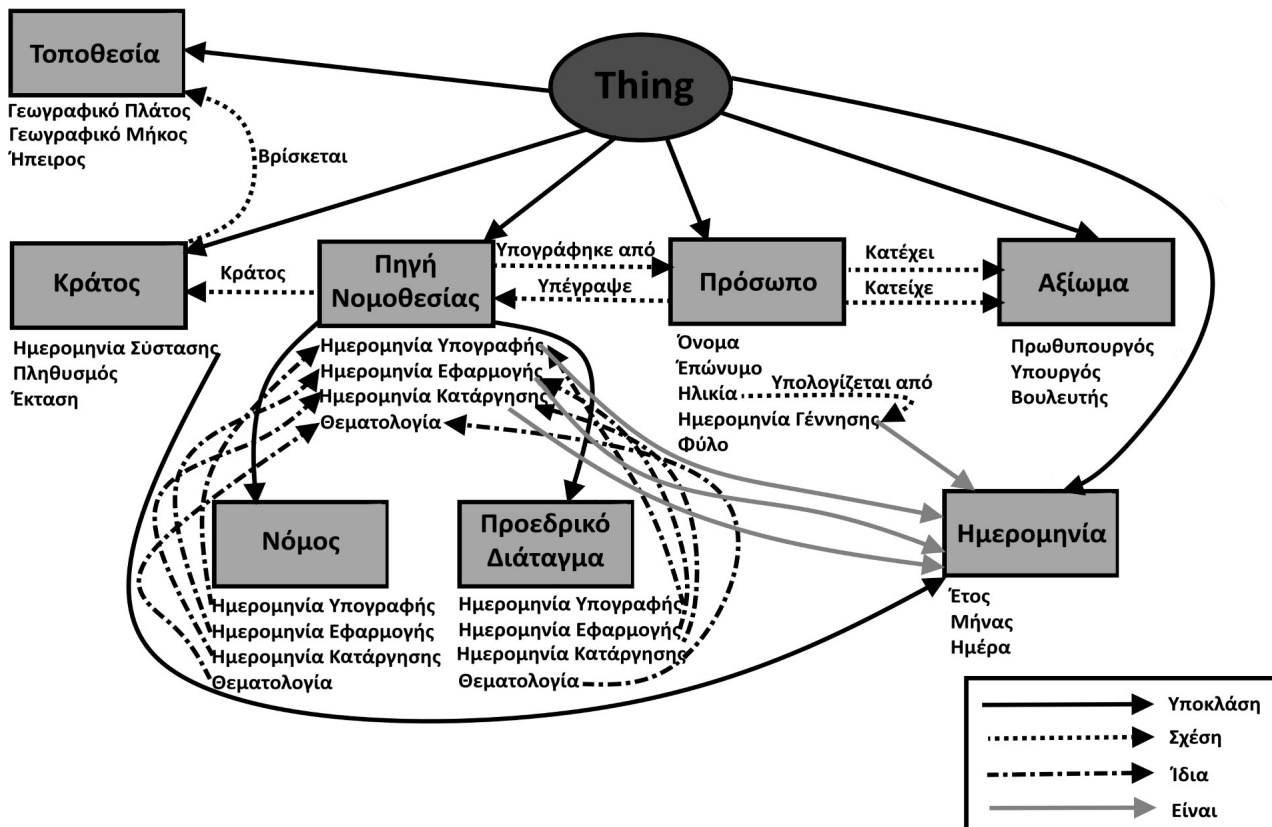
Τα κύρια «συστατικά» μιας οντολογίας, ανεξαρτήτως της γλώσσας στην οποία είναι εκφρασμένη, είναι, συνήθως, τα εξής:

- Κλάσεις (Classes/Concepts): Περιγράφουν έννοιες, σύνολα, συλλογές, είδη αντικειμένων, σελ κ.λ.π. (π.χ. Νόμος).
- Άτομα (Individuals/Instances): Αντικείμενα που περιγράφει η οντολογία, «παραδείγματα» των κλάσεων (π.χ. ο Νόμος 3918/2011).
- Χαρακτηριστικά (Attributes): Ιδιότητες, γνωρίσματα, στοιχεία ή παράμετροι που μπορεί να έχει μια κλάση ή ένα άτομο (π.χ. 02/03/2015 η ημερομηνία έκδοσης (χαρακτηριστικό) του Νόμου 3918/2011 (άτομο)).
- Σχέσεις (Relations): Οι τρόποι με τους οποίους συνδέονται οι κλάσεις και τα άτομα μεταξύ τους (π.χ. ο νόμος (κλάση) εκδίδεται από (σχέση) το Εθνικό Τυπογραφείο (άτομο)).

Άλλα συνήθη «συστατικά» των οντολογιών αποτελούν οι Κανόνες (Rules), οι Περιορισμοί (Restrictions), οι

Όροι Συνάρτησης (Function Terms), τα Γεγονότα (Events), που αποτελούν τις αλλαγές χαρακτηριστικών ή σχέσεων και τα Αξιώματα (Axioms), τα οποία είναι βασικοί ισχυρισμοί και κανόνες για την οντολογία που θεωρούμε εκ των προτέρων πως ισχύουν, καθώς και τα συμπεράσματα που προκύπτουν από αυτά.

Ένα παράδειγμα απεικόνισης μιας οντολογίας που αφορά τη νομοθεσία είναι το παρακάτω:



Εικόνα 4: Παράδειγμα Οντολογίας στο Νομικό Πεδίο

Βλέπουμε πως κάθε «Πηγή Νομοθεσίας» (κλάση) έχει τα χαρακτηριστικά «Ημερομηνία Υπογραφής», «Ημερομηνία Εφαρμογής», «Ημερομηνία Κατάργησης» και «Θεματολογία». Έχει τις υποκλάσεις «Νόμος» και «Προεδρικό Διάταγμα», οι οποίες έχουν τα ίδια χαρακτηριστικά. Συνδέεται μέσω της σχέσης «Κράτος» με τη κλάση «Κράτος», η οποία έχει τα χαρακτηριστικά «Ημερομηνία Σύστασης», «Πληθυσμός» και «Έκταση», ενώ συνδέεται με τη σχέση «Βρίσκεται» με την κλάση «Τοποθεσία», με χαρακτηριστικά «Γεωγραφικό Πλάτος», «Γεωγραφικό Μήκος» και «Ήπειρος». Η «Πηγή Νομοθεσίας» συνδέεται επίσης με την κλάση «Πρόσωπο» μέσω των σχέσεων «Υπογράφηκε από» και «Υπέγραψε», οι οποίες είναι αντίστροφες μεταξύ τους. Η κλάση «Πρόσωπο», με τη σειρά της, έχει τα χαρακτηριστικά «Όνομα», «Έπώνυμο», «Ηλικία», «Ημερομηνία Γέννησης» και «Φύλο», ενώ συνδέεται με τις σχέσεις «Κατέχει» και «Κατείχε» με την κλάση «Αξίωμα», με χαρακτηριστικά «Πρωθυπουργός», «Υπουργός» και «Βουλευτής». Όλες οι ημερομηνίες ανήκουν στην κλάση «Ημερομηνία» με χαρακτηριστικά «Έτος», «Μήνας» και «Ημέρα». Τέλος, όλες οι κλάσεις είναι υποκλάσεις της αρχικής κλάσης «Thing», η οποία είναι η αρχική κλάση σε όλες τις οντολογίες.

Οι οντολογίες κατασκευάζονται από εφαρμογές που ονομάζονται επεξεργαστές οντολογίας (ontology editors) με τη βοήθεια των γλωσσών οντολογιών (ontology languages). Οι περισσότεροι επεξεργαστές οντολογίας χρησιμοποιούν μία ή παραπάνω γλώσσες οντολογιών και έχουν τη δυνατότητα να εξάγουν γνώση και συμπεράσματα από την οντολογία και τα δεδομένα της και να απεικονίσουν οπτικά το μοντέλο γνώσης που εκφράζει η οντολογία. Οι γλώσσες οντολογιών επιτρέπουν την κωδικοποίηση της γνώσης και περιλαμβάνουν κανόνες που βοηθούν στην επεξεργασία της. Πρόκειται συνήθως για γλώσσες δηλωτικού

προγραμματισμού (declarative programming) και βασίζονται είτε σε λογική πρώτου βαθμού (first-order logic) είτε σε περιγραφική λογική (description logic).

Μερικές από τις πιο διαδεδομένες οντολογίες είναι οι εξής:

- Dublin Core (Dublin Core Metadata Element Set - DCMES): Μια απλή οντολογία για την απεικόνιση φυσικών ή ψηφιακών πόρων.
- FOAF (Friend Of A Friend): Οντολογία που περιγράφει ανθρώπους, τις δραστηριότητές τους και τις μεταξύ τους σχέσεις.
- COSMO: Οντολογία με πολύ βασικές έννοιες που έχει ως στόχο με τους όρους της και με κατάλληλες επεκτάσεις, να μπορεί να αναπαραστήσει οποιοδήποτε πεδίο γνώσης.
- Computer Science Ontology: Οντολογία που παράχθηκε αυτόματα και αφορά την επιστήμη των υπολογιστών.
- BFO (Basic Formal Ontology): Ανώτερη οντολογία για την υποστήριξη επιστημονικών ερευνών
- CRM (CICOC Conceptual Reference Model): Οντολογία που αφορά τη πολιτισμική κληρονομιά και τα μουσεία
- FAO Geopolitical Ontology: Οντολογία που περιγράφει, επεξεργάζεται και ανταλλάσσει γεωπολιτικά δεδομένα.
- SBO (Systems Biology Ontology): Οντολογία που αφορά δεδομένα της βιολογίας συστημάτων.

2.1.4 Λεξιλόγια (Vocabularies)

Τα λεξιλόγια χρησιμοποιούνται από τις οντολογίες για να δώσουν όρους στις έννοιες και τις σχέσεις που περιγράφουν το γνωστικό πεδίο της οντολογίας. Στα λεξιλόγια περιλαμβάνονται η ταξινόμηση των όρων αυτών, οι περιορισμοί τους και οι σχέσεις μεταξύ τους. Αποτελούν τη βάση για την παραγωγή νέας γνώσης, καθώς, μέσω των κανόνων τους, μπορούν να ανακαλυφθούν νέες σχέσεις μεταξύ των δεδομένων ή και ασυνέπειες μεταξύ τους με τη χρήση αυτόματων διαδικασιών που ονομάζονται τεχνικές συλλογισμού (inference techniques). Ένα λεξιλόγιο, ανάλογα με την πολυπλοκότητά του και την εφαρμογή στην οποία χρησιμοποιείται, μπορεί να περιλαμβάνει από έναν-δύο όρους (απλά λεξιλόγια) έως και χιλιάδες όρους (σύνθετα λεξιλόγια). Συχνά οι έννοιες της οντολογίας και του λεξιλογίου συγχέονται, με την έννοια της οντολογίας να χρησιμοποιείται περισσότερο σε μεγάλες και επίσημες συλλογές σημασιολογικής γνώσης.

2.1.5 Χώροι Ονομάτων (Namespaces)

Οι χώροι ονομάτων είναι σύνολα ονομάτων που χρησιμοποιούνται για να ορίσουν τις έννοιες ενός συγκεκριμένου γνωστικού πεδίου. Στον Σημασιολογικό Ιστό, οι χώροι ονομάτων εξυπηρετούν στην ομαδοποίηση των ονομάτων αυτών, έτσι ώστε να μην υπάρχουν «συγκρούσεις» σε περίπτωση που το ίδιο όνομα χρησιμοποιείται από διαφορετικές οντολογίες για διαφορετικούς σκοπούς. Για παράδειγμα, το όνομα της κλάσης «Νόμος» έχει διαφορετική έννοια σε μια οντολογία που αφορά τη νομική και σε μια οντολογία που αφορά τη φυσική, πράγμα που σημαίνει ότι οι δυο οντολογίες πρέπει να κάνουν εμφανές μέσω ενός αναγνωριστικού (identifier) ότι πήραν το όνομα αυτό από διαφορετικούς χώρους ονομάτων, έτσι ώστε να μην υπάρξουν λάθη σε περίπτωση που οι αυτές οι οντολογίες χρησιμοποιηθούν μαζί από μια εφαρμογή. Είναι προφανές πως, στον ίδιο χώρο ονομάτων, δεν είναι σωστή τακτική να έχουμε δυο διαφορετικές χρήσεις για το ίδιο όνομα.

2.1.6 Μεταδεδομένα (Metadata)

Τα μεταδεδομένα [19] ορίζονται ως κωδικοποιημένα δεδομένα που δίνουν πληροφορίες για άλλα δεδομένα, χωρίς όμως να περιλαμβάνουν το περιεχόμενό τους. Περιγράφουν χαρακτηριστικά πληροφοριακών οντοτήτων, αποσκοπώντας στην ταύτιση, αναγνώριση, ανακάλυψη, αξιολόγηση και διαχείριση των οντοτήτων που περιγράφονται. Τα δεδομένα τα οποία περιγράφονται από μεταδεδομένα αποτελούν την πηγή. Χάρη στα μεταδεδομένα, η πηγή εμπλουτίζεται με πληροφορία, ενώ παράλληλα η αναζήτησή της γίνεται πιο αποδοτική, καθώς ο εκάστοτε χρήστης απαλλάσσεται από τη δύσκολη και χρονοβόρα χειροκίνητη διαδικασία του φιλτραρίσματος πληροφοριών. Τα μεταδεδομένα συνήθως έχουν δομημένη μορφή και σημασιολογία, ώστε να μπορούν να κατανοούνται και να επεξεργάζονται και από ηλεκτρονικούς υπολογιστές. Ενσωματώνονται στην πηγή και δημοσιεύονται στο διαδίκτυο με αυτήν, ώστε οι εφαρμογές που θα τη χρησιμοποιήσουν να έχουν μια πλήρη περιγραφή της πριν χρειαστεί να προσπελάσουν το περιεχόμενό της.

Υπάρχουν πολλά είδη μεταδεδομένων. Τα περιγραφικά μεταδεδομένα δίνουν περιγραφή της πηγής της πληροφορίας και χρησιμοποιούνται κυρίως για την ανακάλυψή της. Είναι συνήθως δεδομένα όπως ο τίτλος, η περίληψη, ο συγγραφέας κ.λ.π. ενός κειμένου. Τα δομικά μεταδεδομένα δίνουν πληροφορίες για τη δομή, τις σχέσεις και τον τύπο της πηγής, όπως π.χ. ο τρόπος σελιδοποίησης ενός κειμένου ανά ενότητα. Τα διαχειριστικά μεταδεδομένα δίνουν πληροφορίες για τη διαχείριση μιας πηγής, π.χ. το πότε και πώς αυτή δημιουργήθηκε. Τα μεταδεδομένα αναφοράς συνήθως περιγράφουν στατιστικά δεδομένα και δίνουν πληροφορίες για το περιεχόμενο και την ποιότητά τους. Τα στατιστικά μεταδεδομένα περιγράφουν τον τρόπο με τον οποίο συλλέχθηκαν και επεξεργάστηκαν τα στατιστικά δεδομένα της πηγής. Τέλος, τα νομικά μεταδεδομένα αφορούν τον δημιουργό της πηγής, τον κάτοχο των δικαιωμάτων της και την άδεια χρήσης τους, αν αυτή υπάρχει. Ανάλογα με το περιεχόμενό τους, τα μεταδεδομένα μπορεί να περιγράφουν είτε την πηγή, είτε το περιεχόμενό της. Επιπλέον, κάποια μεταδεδομένα είναι μεταβλητά (π.χ. ο αριθμός των δημοσιεύσεων ενός επιστήμονα), ενώ κάποια άλλα είναι αμετάβλητα (π.χ. η ημερομηνία γέννησής του).

Πολλές φορές η γραμμή ανάμεσα σε δεδομένα και μεταδεδομένα είναι λεπτή διότι τα μεταδεδομένα είναι εξ ορισμού δεδομένα. Έτσι, ένα μεταδεδομένο μπορεί ταυτόχρονα να είναι και δεδομένο, ή οι ρόλοι μεταξύ δεδομένων και μεταδεδομένων να αλλάζουν. Για παράδειγμα, σε μια φωτογραφία (δεδομένο), μεταδεδομένα θα μπορούσαν να αποτελούν η τοποθεσία και η ημερομηνία που αυτή τραβήχτηκε, ο τύπος κάμερας που χρησιμοποιήθηκε κ.ο.κ.. Αν όμως έχουμε ένα βίντεο στο οποίο η φωτογραφία αυτή είναι προεπισκόπηση, η ίδια η φωτογραφία μπορεί να αποτελέσει μεταδεδομένο του βίντεο (δεδομένο). Επίσης, είναι δυνατό να υπάρχουν μετα-μεταδεδομένα, δηλαδή μεταδεδομένα που αφορούν άλλα μεταδεδομένα. Στο παράδειγμά μας, ο τύπος κάμερας που χρησιμοποιήθηκε για τη φωτογραφία αποτελεί μετα-μεταδεδομένο του βίντεο (δεδομένο) που έχει την φωτογραφία ως προεπισκόπηση (μεταδεδομένο). Μπορούν, επίσης, να υπάρξουν μετα-μετα-μεταδεδομένα (περιγράφουν μετα-μεταδεδομένα), μετα-μετα-μετα-μεταδεδομένα κ.ο.κ..

Ο Βαθμός Λεπτομέρειας (Granularity) των μεταδεδομένων ποικίλει ανάλογα με το μέγεθος και τους σκοπούς χρήσης της κάθε εφαρμογής. Όσο πιο μεγάλος είναι ο βαθμός λεπτομέρειας, τόσο πιο αναλυτική και δομημένη είναι η πληροφορία και τόσο πιο υψηλού επιπέδου και δυνατοτήτων είναι η επεξεργασία της. Ωστόσο, όσο μεγαλύτερος είναι ο βαθμός λεπτομέρειας των μεταδεδομένων μιας εφαρμογής, τόσο πιο «ακριβό» είναι το χρονικό και οικονομικό κόστος της απόκτησης των πληροφοριών και της δημιουργίας και συντήρησης των αντίστοιχων μεταδεδομένων. Όταν, λοιπόν, σε μια εφαρμογή αυξάνουμε το βαθμό λεπτομέρειας των μεταδεδομένων, θα πρέπει μαζί με τα πλεονεκτήματα να υπολογίζουμε και το αντίστοιχο κόστος, ειδικά αν λάβουμε υπόψη ότι, λόγω της ραγδαίας ανάπτυξης της τεχνολογίας, το σύστημα που αναπτύσσουμε μπορεί σε λίγα χρόνια να θεωρείται απαρχαιωμένο, με αποτέλεσμα όλα μας τα μεταδεδομένα είτε να αχρηστευτούν, εφόσον δε θα υπάρχει πρόσβαση σε αυτά, είτε να χρειαστεί να μεταφερθούν σε κάποιο νέο, πιο μοντέρνο σύστημα, επιφέροντας επιπλέον χρονικό και οικονομικό κόστος.

Τα μεταδεδομένα μπορούν να αποθηκευτούν είτε εσωτερικά, στο ίδιο το αρχείο του δεδομένου που περιγράφουν, είτε εξωτερικά, σε διαφορετικό αρχείο ή πεδίο. Τα εσωτερικά αποθηκευμένα μεταδεδομένα

έχουν το πλεονέκτημα ότι είναι συνεχώς διαθέσιμα μαζί με τα δεδομένα τα οποία αφορούν, με αποτέλεσμα να αλλάζουν άμεσα όταν αλλάζουν και τα δεδομένα, αυξάνοντας τη συνοχή του συστήματος. Το μειονέκτημά τους είναι πως συχνά δημιουργούν πλεονασμούς, ενώ παράλληλα η διαχείρισή τους από ένα κεντρικό σύστημα είναι αδύνατη. Τα εξωτερικά αποθηκευμένα μεταδεδομένα, αντιθέτως, δε δημιουργούν πλεονασμούς και επιτρέπουν πιο αποδοτική αναζήτηση και διαχείριση από το σύστημα που τα χρησιμοποιεί. Ωστόσο, ο διαχωρισμός τους από τα δεδομένα στα οποία αναφέρονται μπορεί να οδηγήσει σε λάθος συγχρονισμό όταν είτε τα δεδομένα είτε τα μεταδεδομένα τους αλλάζουν, δημιουργώντας πλήγμα στη συνοχή του συστήματος.

Ο Διεθνής Οργανισμός Τυποποίησης (International Organisation for Standardisation ή ISO) έχει εκδώσει το πρότυπο ISO/IEC 11179 Metadata Registries (MDR) [20], το οποίο περιγράφει τη διεθνή τυποποίηση των δεδομένων, έτσι ώστε να διευκολύνεται ο ορισμός, η χρήση και ανταλλαγή των μεταδεδομένων από ετερογενή συστήματα. Συνδυάζει τις βασικές αρχές της μοντελοποίησης δεδομένων με δύο αρχές της σημασιολογικής θεωρίας, την σχέση μεταξύ μιας ευρείας έννοιας με πιο συγκεκριμένες έννοιες που σχετίζονται με αυτή, και τη σχέση μιας έννοιας με τους διαφορετικούς τρόπους από τους οποίους αντιπροσωπεύεται. Το πρότυπο ISO/IEC 11179 είναι ουσιαστικά ένας κατάλογος όπου οι συσχετιζόμενοι τύποι δεδομένων ομαδοποιούνται με βάση μια ευρεία έννοια και μια κλάση αντικείμενου, χωρίς όμως να υπάρχει κάποιου τύπου ιεραρχία. Είναι σημαντικό να διευκρινιστεί ότι το πρότυπο ISO/IEC 11179 δεν περιγράφει τα ίδια τα δεδομένα με τη φυσική ή την τεχνική τους μορφή, αλλά ορίζει τη σημασιολογική περιγραφή τους. Παρότι το πρότυπο ISO/IEC 11179 είναι ένα μητρώο δεδομένων, ο σκοπός της δημιουργίας του είναι η υποστήριξη της καταγραφής και της περιγραφής μεταδεδομένων σε οποιοδήποτε είδους εφαρμογή, ώστε αυτά να μπορούν να βρεθούν και να χρησιμοποιηθούν είτε από ανθρώπους είτε από υπολογιστές.

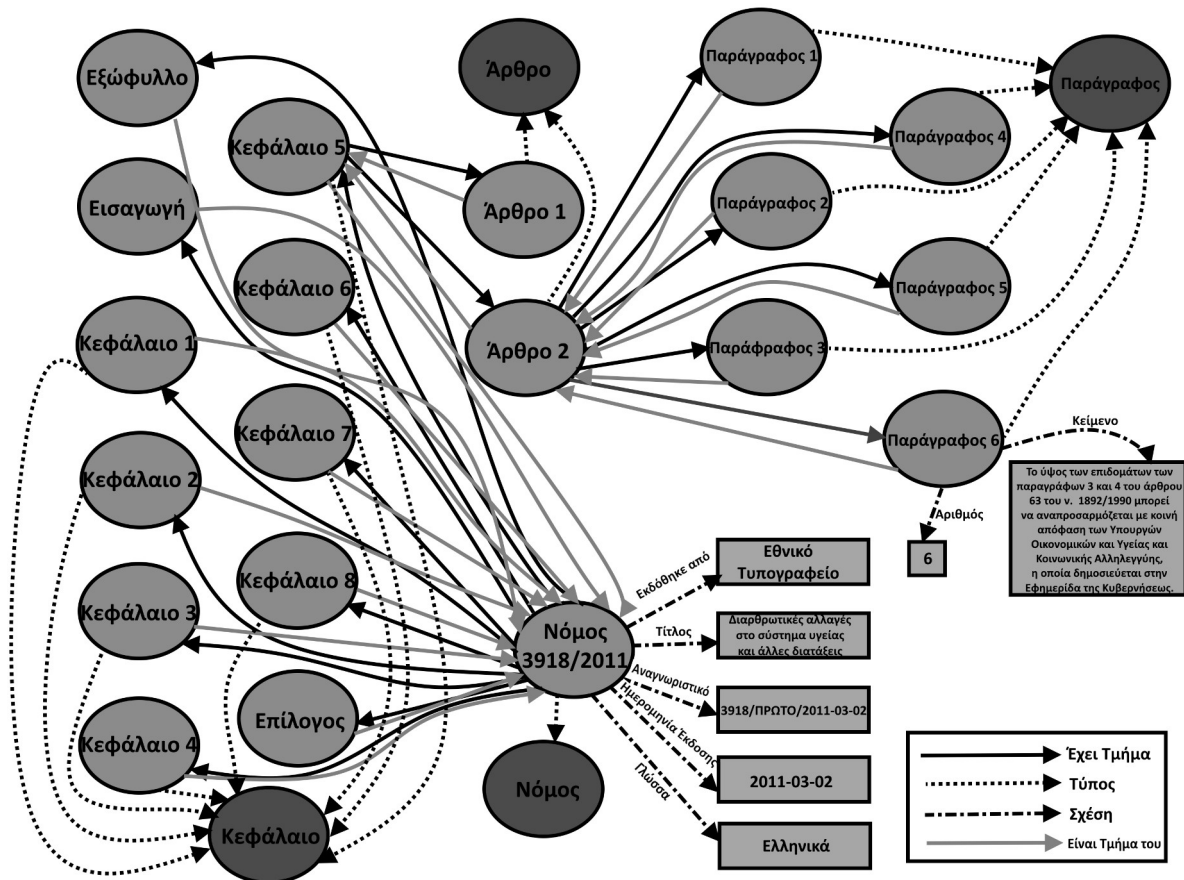
2.2 Τεχνολογίες στο Σημασιολογικό Ιστό

Στην Ενότητα αυτή γίνεται μια συνοπτική παρουσίαση των διαθέσιμων τεχνολογικών εργαλείων που αφορούν την ανάπτυξη εφαρμογών στο Σημασιολογικό Ιστό και τα Διασυνδεδεμένα Δεδομένα.

2.2.1 Το πρότυπο RDF

Το RDF (Resource Description Framework) [21] είναι ένα πρότυπο για την περιγραφή και ανταλλαγή δεδομένων στο Σημασιολογικό Ιστό. Είναι ένα απλό και ευέλικτο μοντέλο, ικανό να περιγράψει οντολογίες και σχέσεις σε σύνθετα πεδία γνώσεων. Τα RDF δεδομένα είναι ένας κατευθυνόμενος γράφος που αποτελείται από «δηλώσεις» σε μορφή τριπλέτας (triple). Μια τριπλέτα αποτελείται από τρία μέρη, το υποκείμενο (subject), την ιδιότητα (predicate ή property) και το αντικείμενο (object) και απεικονίζει μια «δήλωση» γνώσης, π.χ. «Ο νόμος (subject) εκδίδεται από (predicate) το Εθνικό Τυπογραφείο (object)». Οι ιδιότητες χωρίζονται σε resources/object properties, δηλαδή ιδιότητες που σαν αντικείμενο παίρνουν κάποια κλάση ή άτομο της οντολογίας και σε literals/datatype properties, δηλαδή ιδιότητες που έχουν σαν αντικείμενο μια τιμή (π.χ. έναν αριθμό, ένα χρώμα κ.ο.κ.). Τα RDF δεδομένα βρίσκονται σε αποθήκες τριπλετών (triplestores/RDF stores). Οι αποθήκες τριπλετών είναι βάσεις δεδομένων ειδικά διαμορφωμένες για την αποθήκευση και την ανάκτηση δεδομένων σε μορφή τριπλέτας μέσω σημασιολογικών ερωτημάτων (semantic queries).

Ένας RDF γράφος (RDF graph) απεικονίζει τη γνώση μιας οντολογίας και αποτελείται από κόμβους (nodes), οι οποίοι αντιστοιχούν στις κλάσεις και τα άτομα της οντολογίας (είτε αυτά αποτελούν υποκείμενα, είτε αντικείμενα, είτε και τα δύο) και από ακμές (edges), οι οποίες αποτελούν τις ιδιότητες. Ένα παράδειγμα RDF γράφου που αφορά το Νόμο 3918/2011 απεικονίζεται παρακάτω:



Εικόνα 5: Παράδειγμα RDF Γράφου για το Νόμο 3918/2011

Στην παραπάνω εικόνα, το αντικείμενο «Νόμος 3918/2011» έχει τύπο την κλάση «Νόμος». Συνδέεται με τη σχέση «Έχει Τμήμα» με τα αντικείμενα «Εξώφυλλο», «Εισαγωγή», «Επίλογος» και τα Κεφάλαια 1 έως 8, τα οποία έχουν τύπο τη κλάση «Κεφάλαιο». Το «Κεφάλαιο 5» συνδέεται με τη σχέση «έχει Τμήμα» με τα αντικείμενα «Άρθρο 1» και «Άρθρο 2», τα οποία έχουν τύπο τη κλάση «Άρθρο». Το «Άρθρο 2» συνδέεται με τη σχέση «έχει Τμήμα» με τις Παραγράφους 1 έως 6 (αντικείμενα) που έχουν τύπο κλάση «Παράγραφος». Όλες οι Παραγράφοι συνδέονται με το Άρθρο τους με τη σχέση «Είναι Τμήμα του», και το ίδιο συμβαίνει με τα Άρθρα με το Κεφάλαιό τους, και με τα Κεφάλαια με το «Νόμο 3918/2011». Η «Παράγραφος 6» συνδέεται με τις σχέσεις «Αριθμός» και «Κείμενο» με τα αντίστοιχα χαρακτηριστικά της, ενώ ο «Νόμος 3918/2011» συνδέεται με τις σχέσεις «Εκδόθηκε από», «Τίτλος», «Αναγνωριστικό», «Ημερομηνία Έκδοσης» και «Γλώσσα» με τα αντίστοιχα χαρακτηριστικά του.

Κάθε ένα από τα μέρη μιας RDF τριπλέτας μπορεί να αντιπροσωπευτεί με ένα URI (Uniform Resource Identifier - αναλύεται στην Υποενότητα 2.2.2). Τα υποκείμενα και τα αντικείμενα που δεν αντιπροσωπεύονται από κάποιο URI ή (στην περίπτωση των αντικειμένων) δεν έχουν κάποια τιμή ονομάζονται ανώνυμες πηγές (anonymous resources, πιο γνωστές ως blank nodes ή bnodes). Η ύπαρξη ανώνυμων πηγών μπορεί να υπάρξει χρήσιμη σε περιπτώσεις που θέλουμε να αναφερθούμε σε δομές με πολλά μέρη (όπως RDF containers που μπορεί να περιέχουν ομάδες ατόμων) ή σε σύνθετα χαρακτηριστικά με πολλά τμήματα (όπως διευθύνσεις), να χρησιμοποιήσουμε μια ολόκληρη τριπλέτα ως αντικείμενο ή υποκείμενο ή να αποκρύψουμε ευαίσθητα προσωπικά δεδομένα από τρίτους. Ωστόσο, είναι σημαντικό να σημειωθεί πως οι ανώνυμες πηγές αυξάνουν εκθετικά την πολυπλοκότητα ορισμένων εργασιών στους RDF γράφους, ενώ παράλληλα δυσχεραίνουν τον εντοπισμό και την καταγραφή των αλλαγών που γίνονται στη βάση γνώσης.

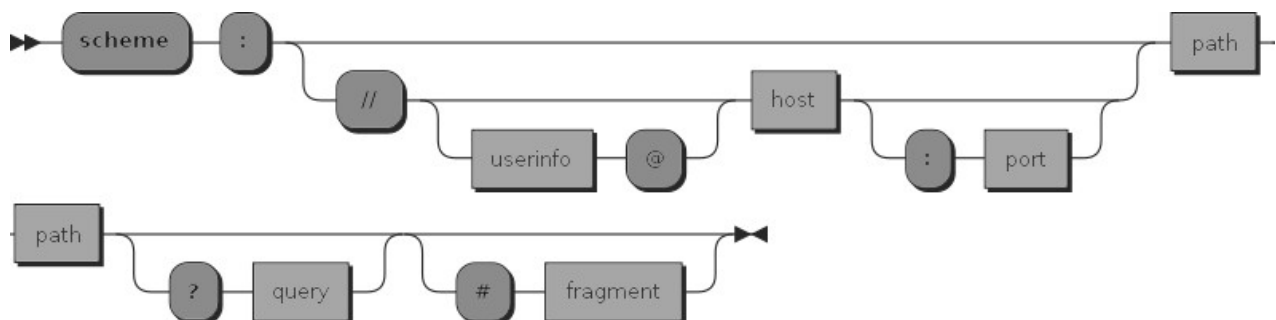
Τα RDF δεδομένα μπορούν να σειριοποιηθούν σε διάφορες μορφές [22]. Οι πιο γνωστές είναι οι εξής:

- Turtle (Terse RDF Triple Language) [23]: Απλή και συμπαγής μορφή τριπλετών, εύκολη στην ανάγνωση για τον άνθρωπο. Επέκτασή της αποτελεί η TriG, που δίνει τη δυνατότητα σειριοποίησης γράφων.
- N3/Notation3 [24]: Ανεπίσημη μορφή παρόμοια με την Turtle, με επιπλέον χαρακτηριστικά, όπως τη δυνατότητα ορισμού συλλογιστικών κανόνων (inference rules).
- N-Triples: Απλή μορφή σειριοποίησης τριπλετών, γεγονός που την κάνει ευκολότερη στην ανάλυση από υπολογιστές και δυσκολότερη στην ανάγνωση από ανθρώπους. Η μορφή N-Quads είναι επέκτασή της και δίνει τη δυνατότητα για μια επιπλέον τιμή στις τριπλέτες.
- RDF/XML [25]: Συντακτικό βασισμένο στη γλώσσα XML (αναλύεται περισσότερο στην Ενότητα 2.2.4). Η μορφή αυτή χρησιμοποιείται στην παρούσα Διπλωματική Εργασία.
- JSON-LD (Javascript Object Notation for Linked Data) [26]: Η JSON αποτελεί μια απλή μορφή αναπαράστασης δεδομένων που χρησιμοποιείται από servers και δικτυακές εφαρμογές (web applications) στην ηλεκτρονική ανταλλαγή δεδομένων. Η μορφή JSON-LD παρέχει τη δυνατότητα στις RDF τριπλέτες να εκφραστούν στη μορφή JSON, διευκολύνοντας, έτσι, το συνδυασμό RDF δεδομένων και JSON εργαλείων και συστημάτων. Με αυτόν τον τρόπο, καθίσταται δυνατή η ενσωμάτωση των JSON συστημάτων του Παγκόσμιου Ιστού στο Σημασιολογικό Ιστό.

2.2.2 Ενιαίος Ταυτοποιητής Πόρων (Uniform Resource Identifier - URI) και Παγκόσμιος Ταυτοποιητής Πόρων (International Resource Identifier - IRI)

Τα URIs [27] είναι μοναδικές ακολουθίες χαρακτήρων που ταυτοποιούν ένα φυσικό ή λογικό πόρο που χρησιμοποιείται από τεχνολογίες του διαδικτύου. Οι πόροι αυτοί μπορούν να είναι οτιδήποτε, όπως φυσικά αντικείμενα, άνθρωποι, τοποθεσίες, ιδέες κ.ο.κ.. Κάποια URIs περιλαμβάνουν τη μέθοδο και τοποθεσία ανάκτησης του πόρου αυτού, είτε πρόκειται για κάποια ιστοσελίδα, είτε για τοποθεσία σε ένα ιδιωτικό δίκτυο ή σύστημα αρχείων. Ο Ενιαίος Εντοπιστής Πόρων (Uniform Resource Locator - URL) παρέχει την τοποθεσία και τη μέθοδο ανάκτησης του πόρου του οποίου το όνομα ταυτοποιεί το URI. Τα URIs που απλά ορίζουν το όνομα του πόρου και δεν περιλαμβάνουν το URL του ονομάζονται Ενιαία Ονόματα Πόρων (Uniform Resource Names - URNs). Τα URIs χρησιμοποιούνται για να ορίσουν πόρους που περιγράφονται από το πρότυπο RDF, όπως έννοιες που περιγράφονται σε οντολογίες δημιουργημένες με τη γλώσσα οντολογίας OWL (αναλύεται στην Ενότητα 2.2.3). Κάθε μια από αυτές τις έννοιες θα έχει το δικό της, μοναδικό URI. Τα URIs που περιλαμβάνουν URLs χρησιμοποιούνται επίσης από περιηγητές (web browsers) για τον εντοπισμό του πόρου στον οποίο αναφέρονται. Για αυτό το λόγο, τα URIs και ειδικότερα τα URLs συχνά συγχέονται με τις έννοιες της ηλεκτρονικής διεύθυνσης (web address) και της ιστοσελίδας (web page), χωρίς όμως να ταυτίζονται με αυτές.

Η συνήθης σύνταξη ενός URI απεικονίζεται παρακάτω:



Εικόνα 6: Συντακτικό Διάγραμμα των URI

Κάθε URI αποτελείται από ένα υποσύνολο επιτρεπόμενων ASCII χαρακτήρων. Κάποιοι από αυτούς είναι δεσμευμένοι (όπως οι / , : , ? , [,] , # , @ και κάποιοι άλλοι ανάλογα με την εφαρμογή) και οι υπόλοιποι είναι αδέσμευτοι και προορίζονται για την περιγραφή του κάθε μέρους του URI. Τα επιμέρους αυτά μέρη είναι τα εξής:

- **scheme**: Το πρωτόκολλο ή το σχήμα που χρησιμοποιείται για τον ορισμό και τον εντοπισμό του πόρου. Κάθε διαφορετικό scheme είναι καταγεγραμμένο στο IANA (Internet Assigned Number Authority) και προσδιορίζει τους συντακτικούς κανόνες που διέπουν το εκάστοτε URI.
- **userinfo, host και port**: Αποτελούν ένα ευρύτερο και προαιρετικό τμήμα του URI που ονομάζεται authority. Το userinfo αποτελείται από το username και πιθανώς το password του εκάστοτε χρήστη, το host είναι είτε η διεύθυνση IP είτε το όνομα του διακομιστή και το port η θύρα που χρησιμοποιείται.
- **path**: Είναι η «διαδρομή» που ακολουθούμε για την εύρεση του συγκεκριμένου πόρου που αναζητούμε. Ξεκινά από τον ευρύτερο «χώρο» (είτε ψηφιακό είτε πραγματικό) και σε κάθε βήμα πηγαίνει σε όλο και ειδικότερο, μέχρι να φτάσουμε στον πόρο. Κάθε βήμα της διαδρομής χωρίζεται από το επόμενο του με "/".
- **query**: Προαιρετικό μέρος του URI που απεικονίζει ένα ερώτημα (query) σε σχέση με τα δεδομένα του πόρου.
- **fragment**: Προαιρετικό μέρος του URI που οδηγεί σε συγκεκριμένο τμήμα του πόρου.

Τα IRIs είναι μια επέκταση των URIs. Τα URIs χρησιμοποιούν μόνο τους αγγλικούς ASCII χαρακτήρες (**US-ASCII**³⁵) για την περιγραφή των πόρων. Αντίθετα, τα IRIs έχουν τη δυνατότητα να χρησιμοποιήσουν πολύ μεγαλύτερη γκάμα χαρακτήρων από διαφορετικές γλώσσες πέραν των αγγλικών, όπως είναι τα κινέζικα, τα κυριλλικά κ.ο.κ.. Στους χαρακτήρες που μπορούν να χρησιμοποιήσουν τα IRIs περιλαμβάνονται οι περισσότεροι χαρακτήρες του Παγκοσμίου Κωδικοποιημένου Σετ Χαρακτήρων (Universal Code Character Set ή UCS ή unicode) που ορίζεται από το πρότυπο **ISO/IEC 10646**³⁶. Για εφαρμογές που δε μπορούν να χρησιμοποιήσουν IRIs, τα IRIs πρέπει να μετατραπούν πρώτα σε URIs με τη βοήθεια ορισμένων τεχνικών που μετατρέπουν κάθε μη αγγλικό χαρακτήρα σε μορφή unicode, δηλαδή σε μια σειρά αγγλικών ASCII χαρακτήρων. Η διαδικασία αυτή είναι αντιστρέψιμη.

Τα IRIs κάνουν πιο εύκολη τη χρήση URIs σε χρήστες που δε γνωρίζουν το λατινικό αλφάβητο και δίνουν περισσότερες δυνατότητες για την ονομασία και ταυτοποίηση των πόρων που περιγράφουν. Ωστόσο, κάθε IRI με χαρακτήρες πέραν των αγγλικών είναι δύσκολο να χρησιμοποιηθεί από χρήστες που δεν διαθέτουν πληκτρολόγιο της αντίστοιχης γλώσσας. Επιπλέον, λόγω της οπτικής ομοιότητας ορισμένων διαφορετικών μεταξύ τους χαρακτήρων, είναι εύκολο ένας χρήστης να χρησιμοποιήσει λάθος IRI είτε λόγω απροσεξίας, είτε λόγω κακόβουλης phishing επίθεσης.

2.2.3 Γλώσσες οντολογίας (Ontology Languages)

Οι γλώσσες οντολογίας [28] είναι γλώσσες προγραμματισμού που χρησιμοποιούνται για τη δημιουργία οντολογιών. Είναι συνήθως γλώσσες δηλωτικού προγραμματισμού και σχεδόν πάντα προέρχονται από γλώσσες πλαισίου (frame languages). Χρησιμοποιούν είτε περιγραφική λογική είτε λογική πρώτου βαθμού. Οι γλώσσες οντολογίας επιτρέπουν την κωδικοποίηση της γνώσης για συγκεκριμένα γνωστικά πεδία και περιλαμβάνουν συλλογιστικούς κανόνες για την επεξεργασία της.

Για να αναπαρίσταται σωστά η γνώση, οι γλώσσες οντολογίας περιλαμβάνουν κλάσεις και υποκλάσεις, ιεραρχία, ιδιότητες και σχέσεις μεταξύ κλάσεων και υποκλάσεων, ιεραρχία και ιδιότητες σχέσεων και συγκεκριμένους περιορισμούς όσον αφορά το τι μπορεί να εκφραστεί. Οι περιορισμοί αυτοί μπορούν να αφορούν το πεδίο τιμών που παίρνουν κάποιες ιδιότητες, το ποιες κλάσεις μπορούν να συσχετιστούν μέσω

35 <http://www.columbia.edu/kermit/ascii.html>

36 [https://standards.iso.org/ittf/PubliclyAvailableStandards/c076835_ISO_IEC_10646_2020\(E\).zip](https://standards.iso.org/ittf/PubliclyAvailableStandards/c076835_ISO_IEC_10646_2020(E).zip)

ποιών σχέσεων και την καρδιναλότητα της οντολογίας (δηλαδή τον τύπο της κάθε σχέσης (1-προς-1, 1-προς-X, X-προς-1, X-προς-X, κ.ο.κ.) ή τον αριθμό των σχέσεων ενός είδους που μπορεί να έχει μία κλάση).

Μια γλώσσα οντολογίας οφείλει να έχει καλά ορισμένο συντακτικό και σημασιολογία, αποδοτική υποστήριξη συλλογισμού, επαρκή εκφραστική δύναμη και ευκολία έκφρασης [1]. Αυτά τα χαρακτηριστικά είναι ζωτικής σημασίας, διότι εξασφαλίζουν πως το νόημα της γνώσης εκφράζεται με ακρίβεια, δεν είναι διφορούμενο και «κατανοείται» με τον ίδιο τρόπο από όλους τους διαφορετικούς ανθρώπους και όλα τα διαφορετικά συστήματα. Ιδιαίτερος σημαντική είναι και η ισορροπία μεταξύ εκφραστικής δύναμης και αποδοτικότητας συλλογισμού. Μια γλώσσα οντολογίας πρέπει να είναι αρκετά πλούσια ώστε να μπορεί να εκφράσει εκτεταμένα πεδία γνώσης, όμως πρέπει να προσέξουμε ώστε ο πλούτος αυτός να μην οδηγεί σε μη αποδοτικό συλλογισμό ή σε αδυναμία υπολογισιμότητας από το σύστημα που χρησιμοποιούμε.

Μερικές από τις σημαντικότερες γλώσσες οντολογίας είναι οι εξής:

- RDF Schema (Resource Description Framework Schema, επίσης γνωστό και ως RDFS, RDF(S), RDF-S ή RDF/S) [29]: Πρόκειται για μια γλώσσα σήμανσης που αποτελείται από κλάσεις με συγκεκριμένες ιδιότητες και ακολουθεί το μοντέλο αναπαράστασης γνώσης RDF. Χρησιμοποιείται για τη δημιουργία RDF πόρων. Σημαντικά RDF λεξιλόγια που χρησιμοποιούν την RDFS είναι το Dublin Core, το FOAF (Friend of a Friend), το SKOS (Simple Knowledge Organisation System), το οποίο απεικονίζει ελεγχόμενα λεξικά και ταξονομίες (taxonomies), το Schema.org (μέχρι το 2020) και το Library Of Congress.

Η πρώτη έκδοση της RDFS δημοσιεύτηκε από την W3C τον Απρίλιο του 1998, ενώ η πιο πρόσφατη τον Φεβρουάριο του 2014. Παρότι, όπως αναφέρθηκε, χρησιμοποιείται σε αρκετά και σημαντικά λεξιλόγια, από τις αρχές του 21ου αιώνα δημιουργήθηκε η ανάγκη για γλώσσες οντολογίας με περισσότερες δυνατότητες. Σήμερα, η RDFS χρησιμοποιείται κυρίως από χρήστες που χρειάζονται μια ιεραρχία ταξινόμησης με τυποποίηση ιδιοτήτων και με δυνατότητα μοντελοποίησης μεταδεδομένων.

- DAML+OIL [30]: Η OIL (Ontology Inference Layer ή Ontology Interchange Language) είναι μια από τις πρώτες γλώσσες οντολογίας που αναπτύχθηκαν. Χρησιμοποιεί περιγραφική λογική, είναι επέκταση της RDFS και συμβατή με αυτήν και βασίστηκε στην XML. Το 2001 συγχωνεύθηκε με την DAML (DARPA Agent Markup Language) και με στοιχεία της SHOE (Simple HTML Ontology Extensions), δημιουργώντας τη γλώσσα DAML+OIL. Η γλώσσα αυτή χρησιμοποιεί περιγραφική λογική και τυπική σημασιολογία (formal semantics) και αποτέλεσε τη βάση για την ανάπτυξη της OWL.

- OWL (Web Ontology Language) [31]: Πρόκειται για μια οικογένεια γλωσσών οντολογίας, βασισμένη στο μοντέλο RDF και με αρκετές επιρροές από την DAML+OIL. Αποτελεί και αυτή γλώσσα σήμανσης. Άρχισε να αναπτύσσεται στις αρχές της δεκαετίας του 2000, όταν η ανάγκη για μια γλώσσα πιο εκφραστική από την RDFS έγινε εμφανής. Οι γλώσσες OWL είναι γλώσσες σήμανσης, ενώ χαρακτηριστικό τους είναι η τυπική σημασιολογία, δηλαδή το γεγονός πως διαθέτουν ένα λεπτομερώς ορισμένο και αυστηρό σύστημα με μαθηματικές δομές για να προσδιορίζεται επαρκώς και με ακρίβεια το νόημα των κατασκευασμάτων που δημιουργούμε με αυτές. Χρησιμοποιούν την υπόθεση ανοιχτού κόσμου (open world assumption), σύμφωνα με την οποία η έλλειψη δεδομένων για την απόδειξη πως μια υπόθεση αληθεύει δε σημαίνει αυτόματα πως η υπόθεση αυτή είναι ψευδής.

Οι γλώσσες της οικογένειας αυτής είναι οι παρακάτω:

- i. OWL Lite: Ένας από τους σημαντικότερους περιορισμούς της RDFS είναι η αδυναμία του χρήστη να ορίσει ισότητα ή ανισότητα μεταξύ των κλάσεων ή των ατόμων μιας οντολογίας. Στην OWL Lite η δυνατότητα αυτή υποστηρίζεται, ενώ τα άτομα που δημιουργούνται δεν θεωρούνται αυτόματα διαφορετικά μεταξύ τους αν έχουν διαφορετικό όνομα, όπως γίνεται σε άλλες γλώσσες. Αντίστοιχες δομές υπάρχουν και για τις ιδιότητες της οντολογίας, καθώς μπορούν να ορίζονται αντίστροφες (inverse), συμμετρικές (symmetric) και διαδοχικές (transitive) ιδιότητες. Παράλληλα, υπάρχει η δυνατότητα ο χρήστης να θέσει

περιορισμούς στο πεδίο ορισμού (domain) και στο πεδίο τιμών (range) μιας ιδιότητας, όπως επίσης και να ορίσει αν μια ιδιότητα κάποιας κλάσης είναι υποχρεωτικό να έχει τιμή ή όχι. Η OWL Lite υποστηρίζει μόνο καρδιναλότητα 0 και 1, πρόβλημα που λύνει η OWL DL. Ωστόσο, στην πράξη, με τον κατάλληλο συνδυασμό των χαρακτηριστικών της και με κατάλληλες μακροεντολές (macros), μπορεί και η OWL Lite να υποστηρίζει καρδιναλότητα πάνω από 1 χάρη στην εκφραστικότητα της, δίνοντάς της τη δυνατότητα να εκφράσει σχεδόν όλες τις δομές που μπορεί να εκφράσει και η OWL DL με πιο σύνθετο, όμως, τρόπο. Το γεγονός αυτό κάνει τα εργαλεία της OWL Lite να αναπτύσσονται με την ίδια δυσκολία που αναπτύσσονται αυτά της OWL DL, με αποτέλεσμα να προτιμάται η OWL DL και η OWL Lite να χρησιμοποιείται σπάνια.

ii. OWL DL: Όπως αναφέρθηκε, η OWL DL είναι ένα βήμα μπροστά από την OWL Lite, καθώς περιλαμβάνει επιπλέον δομές που απλοποιούν την ανάπτυξη μιας οντολογίας. Μερικές από τις δομές αυτές είναι οι ασυμπτωτικές (disjoint) κλάσεις, η δυνατότητα χρήσης αυθαίρετων λογικών αλγεβρικών εκφράσεων κατά τον ορισμό αξιωμάτων και η δυνατότητα καρδιναλότητας πάνω από 1. Έχει μεγαλύτερη εκφραστικότητα και πληρότητα από την OWL Lite, ενώ, παράλληλα, περιλαμβάνει συλλογιστικούς αλγορίθμους. Στην OWL DL περιλαμβάνονται όλες οι διαθέσιμες δομές της OWL, αλλά έχουν ορισμένους περιορισμούς ώστε να μπορεί να εφαρμοστεί περιγραφική λογική και να είναι η γλώσσα επιλύσιμη (decidable). Πολλοί από τους περιορισμούς αυτούς δεν υπάρχουν στην OWL Full.

iii. OWL Full: Σε αντίθεση με την OWL Lite και την OWL DL, που χρησιμοποιούν περιγραφικές λογικές, η OWL Full δημιουργήθηκε για να είναι συμβατή με την RDFS. Στην OWL Full υπάρχει μεγαλύτερη ελευθερία στην εκφραστικότητα και επιτρέπονται δηλώσεις που απαγορεύει η OWL DL, ενώ οι δομές της OWL Full αποτελούν και οι ίδιες σημασιολογικά αντικείμενα, κάτι που δε συμβαίνει στις δύο προηγούμενες γλώσσες. Έτσι, μία κλάση μπορεί να εξισωθεί με ένα άτομο, να είναι instance του εαυτού της και να συνδέεται με μια άλλη κλάση με παραπάνω από μια σχέσεις. Η OWL Full εμπλουτίζει τη σημασιολογία RDF και RDFS λεξιλογίων, αποτελώντας, ουσιαστικά, σημασιολογική επέκταση του μοντέλου RDF. Σε αντίθεση με τις δύο προηγούμενες, η OWL Full είναι μη επιλύσιμη (undecidable) εξαιτίας της τόσο μεγάλης ελευθερίας εκφραστικότητας της, με αποτέλεσμα να είναι αδύνατο να αναπτυχθούν συλλογιστικές μηχανές (reasoners) για αυτήν.

Εύκολα συμπεραίνουμε πως κάθε οντολογία ή συμπέρασμα της OWL Lite είναι οντολογία ή συμπέρασμα της OWL DL αντίστοιχα. Ομοίως, κάθε οντολογία ή συμπέρασμα της OWL DL είναι οντολογία ή συμπέρασμα της OWL Full αντίστοιχα. Το αντίστροφο δεν ισχύει, καθώς κάθε μια από τις παραπάνω γλώσσες είναι, ουσιαστικά, επέκταση της προηγούμενης.

• OWL2 [32]: Η OWL2 είναι η εξέλιξη της OWL. Έχει λιγότερους περιορισμούς από την OWL DL, επιτρέποντας μεγαλύτερη εκφραστικότητα. Περιλαμβάνει επιπλέον χαρακτηριστικά και δομές όπως:

i. Κλειδιά (Keys): Αποτελούν έναν συνδυασμό ιδιοτήτων ενός ατόμου. Ο συνδυασμός των τιμών των ιδιοτήτων αυτών ορίζει μοναδικά το άτομο. Με τη βοήθεια των κλειδιών υπάρχει η δυνατότητα να ορίσουμε τα στοιχεία που κάνουν μοναδικά τα άτομα μιας κλάσης και επιτυγχάνεται αξιοπιστία δεδομένων.

ii. Αλυσίδες Ιδιοτήτων (Property Chains): Με τη βοήθειά τους μπορούμε να συμπεράνουμε νέες σχέσεις μεταξύ των κλάσεων ή των ατόμων μιας οντολογίας, παράγοντας, έτσι, νέα γνώση.

iii. Περισσότερους Τύπους Δεδομένων (Datatypes) και Πεδία Τιμών

iv. Εμπλουτισμένους Περιορισμούς Καρδιναλότητας: Οι περιορισμοί καρδιναλότητας δεν αφορούν πλέον μόνο την ποσότητα των ατόμων ή κλάσεων που μπορούν να συνδέονται με μια συγκεκριμένη σχέση. Υπάρχει η δυνατότητα να τεθούν και νέου είδους περιορισμοί, όπως, για παράδειγμα, ότι οι Νόμοι (κλάση), υπογράφονται (ιδιότητα) μόνο από (περιορισμός) Υπουργούς (κλάση) οι οποίοι έκαναν τη θητεία τους την περίοδο (χαρακτηριστικό της κλάσης Υπουργός) που εκδόθηκε ο Νόμος (χαρακτηριστικό της κλάσης Νόμος).

v. Ασύμμετρες (Asymmetric), Αντανακλαστικές (Reflexive) και Ασύνδετες (Disjoint) Ιδιότητες

vi. Βελτιωμένες Δυνατότητες Σχολιασμού (Enhanced Annotation Capabilities): Προσφέρουν περισσότερες δυνατότητες για τον σχολιασμό κλάσεων, ατόμων, χαρακτηριστικών και ιδιοτήτων μιας οντολογίας,

εμπλουτίζοντάς την με μεγαλύτερη ποσότητα και ποικιλία μεταδεδομένων.

Η OWL2 χωρίζεται σε τρεις ξεχωριστές «γλώσσες», τα λεγόμενα προφίλ (OWL2 Profiles). Κάθε προφίλ είναι ένα υποσύνολο από δομές της OWL2, και όλα τους είναι πιο περιοριστικά από την OWL DL. Έτσι, ενώ μειώνεται η εκφραστικότητα, κάθε προφίλ είναι πιο αποδοτικό στον σκοπό για τον οποίο έχει δημιουργηθεί. Τα προφίλ αυτά είναι τα εξής:

- i. **OWL2 EL**: Επιτρέπει συλλογιστικούς αλγορίθμους πολυωνυμικού χρόνου και χρησιμοποιείται κυρίως σε οντολογίες μεγάλου όγκου, όπου δεν απαιτείται μεγάλη εκφραστικότητα. Έτσι, η εκφραστικότητα που χάνεται μετατρέπεται σε αποδοτικότητα, η οποία έχει μεγαλύτερη σημασία στην παραπάνω περίπτωση.
- ii. **OWL2 QL**: Επιτρέπει να απαντώνται ερωτήματα (queries) και δίνει άμεση πρόσβαση στα δεδομένα των οντολογιών στις οποίες χρησιμοποιείται. Συνήθως πρόκειται για σχετικά μικρές οντολογίες οι οποίες καλούνται να διαχειριστούν μεγάλο πλήθος ατόμων.
- iii. **OWL2 RL**: Επιτρέπει την εφαρμογή συλλογιστικών αλγορίθμων πολυωνυμικού χρόνου σε δεδομένα με μορφή RDF τριπλετών. Χρησιμοποιείται συνήθως σε μικρές οντολογίες με μεγάλο αριθμό ατόμων.

2.2.4 Extensible Markup Language (XML)

Η XML [33] είναι γλώσσα σήμανσης και τυποποιημένη μορφή αρχείου για την αποθήκευση, την επεξεργασία και τον διαμοιρασμό δεδομένων μεταξύ διαφορετικών εφαρμογών. Ορίζει ένα σύνολο κανόνων έτσι ώστε τα δεδομένα να κωδικοποιούνται σε αρχεία τα οποία κατανοούνται με τον ίδιο τρόπο και από μηχανές και από ανθρώπους. Η XML είναι σχεδιασμένη ώστε να δίνει έμφαση στην απλότητα, τη γενικότητα και τη χρησιμότητα στο διαδίκτυο. Τα δεδομένα αποθηκεύονται σε μορφή κειμένου, ενώ παράλληλα υποστηρίζει τη χρήση όλων των unicode χαρακτήρων. Χρησιμοποιείται ευρέως για την αναπαράσταση αυθαίρετων δομών δεδομένων σε ποικίλες διαδικτυακές υπηρεσίες.

Ένα XML αρχείο αποτελείται από μια σειρά χαρακτήρων. Κάθε χαρακτήρας αποτελεί είτε τη σήμανση είτε το περιεχόμενο της πληροφορίας. Η σήμανση ξεχωρίζει από το περιεχόμενο κυρίως με τη με τη βοήθεια των ετικετών (tags). Κάθε ετικέτα ξεκινά (ή «ανοίγει») με το χαρακτήρα < και τελειώνει (ή «κλείνει») με τον χαρακτήρα >». Μέσα στην ετικέτα περιέχεται η σήμανση, ενώ εκτός αυτής βρίσκεται το περιεχόμενο. Η XML χρησιμοποιεί τις ετικέτες για την κατηγοριοποίηση και την οργάνωση της πληροφορίας. Υπάρχουν τριών ειδών ετικέτες: οι ετικέτες-αρχής (start-tags) με τη μορφή <tag>, οι ετικέτες-τέλους (end-tags) με τη μορφή </tag> και οι ετικέτες-χωρίς-περιεχόμενο (empty-element tags) με τη μορφή <tag/>.

Ένα στοιχείο (element) ξεκινά με μια ετικέτα-αρχής και τελειώνει με μια ετικέτα-τέλους με το ίδιο όνομα. Μεταξύ της ετικέτας-αρχής και της ετικέτας-τέλους βρίσκεται το περιεχόμενο του στοιχείου, το οποίο μπορεί να περιλαμβάνει σήμανση και άλλα στοιχεία (που ονομάζονται στοιχεία-παιδιά (child-elements)). Ένα στοιχείο μπορεί επίσης να είναι άδειο ή να αποτελείται μόνο από μια ετικέτα-χωρίς-περιεχόμενο.

Άλλη μια ιδιότητα της σήμανσης είναι τα χαρακτηριστικά (attributes). Αυτά εμπεριέχονται μέσα στην ετικέτα και είναι ένα ζεύγος ονόματος και τιμής (name/value pair). Η μορφή που παίρνει η ετικέτα για δυο χαρακτηριστικά είναι η <tag name1=value1 name2=value2>. Κάθε διαφορετικό χαρακτηριστικό μπορεί να εμφανίζεται το πολύ μια φορά στην ίδια ετικέτα και να παίρνει μόνο μια τιμή.

Τα αρχεία XML «διαβάζονται» από επεξεργαστές που είναι γνωστοί ως XML parsers. Οι XML parsers αναλύουν τη σήμανση ανάλογα με τις απαιτήσεις της εκάστοτε εφαρμογής και μεταφέρουν στην εφαρμογή αυτή τα δεδομένα (δηλαδή τα κομμάτια από το περιεχόμενο του αρχείου) που χρειάζεται. Υπάρχουν κανόνες σχετικά με το τι επιτρέπεται να κάνουν οι XML parsers και τι όχι, όμως οι κανόνες αυτοί δεν αφορούν την εφαρμογή και τη χρήση που αυτή θα κάνει στα δεδομένα που ο επεξεργαστής θα της δώσει.

Στο Σημασιολογικό Ιστό, η XML χρησιμοποιείται για τη δημιουργία αρχείων που περιγράφουν RDF

γράφους δεδομένων. Το συντακτικό των αρχείων αυτών ονομάζεται RDF/XML και έχει οριστεί από την W3C. Άλλα εργαλεία της XML που χρησιμοποιούνται στο Σημασιολογικό Ιστό είναι τα εξής:

- XML Χώροι Ονομάτων (XML Namespaces) [34]: Οι χώροι ονομάτων χρησιμοποιούνται για την αποφυγή συγκρούσεων σε περίπτωση που δυο γνωστικά πεδία χρησιμοποιούν την ίδια λέξη για να ορίσουν διαφορετικές έννοιες. Στην αρχή των XML αρχείων δηλώνεται ο προκαθορισμένος χώρος ονομάτων (default namespace) και οι υπόλοιποι χώροι ονομάτων που θα χρησιμοποιηθούν μαζί με το αναγνωριστικό πρόθεμα (prefix) του καθενός με τη βοήθεια του xmlns χαρακτηριστικού της XML. Αν ένα στοιχείο (ή κάποιο στοιχείο-γονιός του) του αρχείου XML δεν έχει μπροστά του κάποιο πρόθεμα, θεωρείται πως ανήκει στον προκαθορισμένο χώρο ονομάτων. Αν αυτός δεν έχει οριστεί, τότε το στοιχείο αυτό δεν ανήκει σε κανέναν χώρο ονομάτων. Ο προκαθορισμένος χώρος ονομάτων δεν ισχύει και για τα χαρακτηριστικά. Έτσι, ένα χαρακτηριστικό είτε έχει μπροστά του πρόθεμα που δηλώνει τον χώρο ονομάτων από τον οποίο προέρχεται, είτε δεν ανήκει σε κανένα χώρο ονομάτων. Οι χώροι ονομάτων στα XML αρχεία συνήθως είναι URIs ή IRIs, όμως οι περισσότεροι XML parsers δεν απαιτούν τα ονόματα των χώρων ονομάτων να ακολουθούν τη σημασιολογία των URIs/IRIs, παρά μόνο το συντακτικό και τη μορφή τους.
- Extensible Stylesheet Language Transformations (XSLT) [35]: Η XSLT είναι δηλωτική γλώσσα προγραμματισμού που χρησιμοποιείται για την μετατροπή XML αρχείων σε άλλα XML ή άλλου τύπου (π.χ. HTML) αρχεία. Οι κανόνες της μετατροπής αυτής δηλώνονται σε XSLT τύπου αρχεία τα οποία ονομάζονται stylesheets και έχουν XML σύνταξη. Τα αρχεία εισόδου πάντα διατηρούνται, με την XSLT να δημιουργεί τα αρχεία εξόδου που ορίζει το stylesheet ή τα stylesheets που χρησιμοποιούνται κάθε φορά. Στην περίπτωση του Σημασιολογικού Ιστού, XML αρχεία δεδομένων μπορούν να μετατραπούν σε XML αρχεία με συντακτικό RDF/XML, ή τα ήδη υπάρχοντα XML αρχεία με RDF/XML συντακτικό μπορούν να μετατραπούν σε άλλης μορφής αρχεία για την παρουσίαση των δεδομένων στο χρήστη.
- XML Path Language (XPath) [36]: Αποτελεί μια γλώσσα εκφράσεων (expression language) για την υποστήριξη μετατροπών και ερωτημάτων σε XML αρχεία. Χρησιμοποιείται για τον υπολογισμό τιμών (αριθμητικών, λογικών (boolean) ή συμβολοσειρών (strings)) που προκύπτουν από τα περιεχόμενα ενός XML αρχείου. Βασίζεται στην αναπαράσταση του XML αρχείου ως δέντρο και δίνει τη δυνατότητα να επιλέξουμε όποιους κόμβους-στοιχεία χρειαζόμαστε, με βάση τους κανόνες και τους περιορισμούς που θέτουμε. Εκφράσεις XPath (XPath expressions ή απλά XPaths) χρησιμοποιούνται από πολλές εφαρμογές, όπως επίσης και από την XSLT.

2.2.5 Γλώσσες Ερωτημάτων στο Σημασιολογικό Ιστό

Οι γλώσσες ερωτημάτων (query languages/data query languages/database query languages/DQLs) χρησιμοποιούνται για την ανάκτηση πληροφοριών από βάσεις δεδομένων και πληροφοριακά συστήματα με βάση κριτήρια τα οποία θέτει ο εκάστοτε χρήστης. Για το Σημασιολογικό Ιστό έχουν αναπτυχθεί ειδικές γλώσσες ερωτημάτων. Οι γλώσσες αυτές, σε αντίθεση με τις συνήθεις γλώσσες ερωτημάτων, δεν είναι σχεδιασμένες για να διατρέχουν δεδομένα μόνο από μία βάση δεδομένων, αλλά για να μπορούν να ανακτήσουν και να συνδυάσουν δεδομένα από πολλές διαφορετικές πηγές στο Σημασιολογικό αλλά και στον Παγκόσμιο Ιστό. Οι πιο σημαντικές είναι οι εξής:

- XQuery [37]: Αποτελεί γλώσσα ερωτημάτων χωρίς παρενέργειες (side effect-free) για δομημένα ή αδόμητα δεδομένα σε μορφή XML. Βασίζεται στο Μοντέλο Δεδομένων XQuery και XPath (XQuery and XPath Data Model ή XDM), το οποίο αναπαριστά τις πληροφορίες σε μορφή δέντρου. Χρησιμοποιεί XPath εκφράσεις και μπορεί να δημιουργήσει νέα XML αρχεία όπως και η XSLT. Η XSLT είναι πιο ισχυρή και προσφέρει περισσότερες δυνατότητες σε αυτόν τον τομέα, δεδομένου ότι αναπτύχθηκε ειδικά για τη μετατροπή XML αρχείων. Η XQuery, ωστόσο, είναι εκφραστικά απλούστερη από την XSLT, γεγονός που την κάνει πιο εύχρηστη και της δίνει τη δυνατότητα να χρησιμοποιείται και από άλλες γλώσσες σε μορφή εμφωλευμένων XQuery ερωτημάτων.

- **SPARQL (Simple Protocol And RDF Query Language)** [38]: Η SPARQL είναι μια σημασιολογική γλώσσα ερωτημάτων σχεδιασμένη για χρήση σε δεδομένα RDF μορφής και αποτελεί μια από τις τεχνολογίες-κλειδιά του Σημασιολογικού Ιστού. Εφόσον τα δεδομένα είναι σε RDF μορφή, η βάση δεδομένων είναι ουσιαστικά ένα σύνολο από τριπλέτες της μορφής υποκείμενο-ιδιότητα-αντικείμενο. Με τη SPARQL θέτουμε στη βάση ερωτήματα χρησιμοποιώντας μοτίβα της μορφής αυτής, όπου κάθε στοιχείο της τριπλέτας μπορεί να αποτελεί μια άγνωστη μεταβλητή. Μέσω αντιστοίχισης των μοτίβων αυτών (pattern matching) με τα δεδομένα-τριπλέτες της βάσης, η SPARQL θέτει τις τιμές που πρέπει στις άγνωστες μεταβλητές και μας δίνει όσα δεδομένα ταιριάζουν με το μοτίβο του ερωτήματος. Ουσιαστικά, τα μοτίβα που χρησιμοποιούμε στο SPARQL ερώτημα αντικατοπτρίζουν τα κριτήρια και τους περιορισμούς που θέλουμε να έχουν τα δεδομένα ζητάμε.

Ένα ερώτημα SPARQL μπορεί να περιλαμβάνει λογικούς και αριθμητικούς τελεστές, συζεύξεις/τομές (conjunctions/intersections), διαζεύξεις/ενώσεις (disjunctions/unions) και προαιρετικά μοτίβα. Τα προαιρετικά μοτίβα χρησιμεύουν ώστε να μην αποκλειστούν πιθανές λύσεις του SPARQL ερωτήματος στην περίπτωση που δεν διαθέτουν δεδομένα για κάποιο από τα μοτίβα του ερωτήματος. Τα προαιρετικά μοτίβα είναι απαραίτητα διότι δεν είναι όλα τα RDF δεδομένα δομημένα επαρκώς, με αποτέλεσμα συχνά κάποιες ιδιότητες ή χαρακτηριστικά τους να λείπουν από τη βάση δεδομένων.

Η SPARQL είναι καταλληλότερη της XQuery για χρήση σε RDF δεδομένα διότι η δεύτερη γίνεται πολύ συνθετότερη όταν πρέπει να προσπελαστούν πολλές σχέσεις, φαινόμενο αρκετά συχνό στο Σημασιολογικό Ιστό. Επιπλέον, η SPARQL προσφέρει τη δυνατότητα να δημιουργήσει νέους RDF γράφους με τα δεδομένα εξόδου, όπως επίσης και να απαντήσει και στο αν υπάρχουν δεδομένα τα οποία ταιριάζουν στα μοτίβα που ορίζει το ερώτημα. Η SPARQL μπορεί επίσης να αντλήσει δεδομένα από πολλές διαφορετικές πηγές, χωρίς να είναι απαραίτητο τα δεδομένα τους να είναι ταξινομημένα με ομοιόμορφο τρόπο.

Η SPARQL χρησιμοποιείται από πολλές εφαρμογές οι οποίες δίνουν τη δυνατότητα στο χρήστη να θέσει SPARQL ερωτήματα στα δεδομένα τους μέσω SPARQL endpoints. Επιπλέον, υπάρχουν εργαλεία τα οποία μετατρέπουν SPARQL ερωτήματα σε ερωτήματα XQuery ή SQL.

Οι επεκτάσεις της SPARQL περιλαμβάνουν τη SPARQL/Update (SPARUL), η οποία δίνει στο χρήστη τη δυνατότητα να διαγράψει, να εισάγει ή να επεξεργαστεί τα δεδομένα του RDF γράφου, τη GeoSPARQL, η οποία εξειδικεύεται σε γεωχωρικά δεδομένα και τη SPARQL MM, η οποία εξειδικεύεται σε δεδομένα με τη μορφή πολυμέσων.

- **RQL (RDF Query Language)** [39]: Η RQL χρησιμοποιείται σε RDF μορφής δεδομένα, λαμβάνοντας υπόψη την οντολογία με την οποία είναι συνδεδεμένα. Σε αντίθεση με την SPARQL, δεν αντιμετωπίζει τα δεδομένα αυτά ως τριπλέτες, ενώ έχει και τη δυνατότητα να θέσει ερωτήματα και στο οντολογικό σχήμα στο οποίο ανήκουν αυτά τα δεδομένα. Είναι πιο εκφραστική από την SPARQL, γεγονός όμως που την κάνει να έχει υπερβολικά πολλά χαρακτηριστικά και ασυνήθιστο συντακτικό, γεγονός που απαιτεί περισσότερη εμπειρία κατά τη χρήση της. Ακόμα, τα RDF δεδομένα για τα οποία η RQL έχει σχεδιαστεί πρέπει να έχουν κάποιους επιπλέον περιορισμούς σε σχέση με τα συνήθη RDF δεδομένα, γεγονός που την κάνει ακατάλληλη για πιο γενικευμένη χρήση, παρά το γεγονός ότι προσφέρει περισσότερες δυνατότητες.

2.2.6 Εργαλεία Δημοσίευσης Διασυνδεδεμένων Δεδομένων

Για τη δημοσίευση των Διασυνδεδεμένων Δεδομένων έχουν αναπτυχθεί ορισμένα εργαλεία τα οποία παρέχουν τις κατάλληλες υποδομές για τη φιλοξενία συνόλων RDF δεδομένων (RDF datasets). Τα εργαλεία αυτά είτε παρέχουν τα περιεχόμενα RDF συνόλων δεδομένων σε μορφή Διασυνδεδεμένων Δεδομένων στο διαδίκτυο, είτε παρέχουν δεδομένα από παλαιότερου τύπου πηγές δεδομένων σε μορφή Διασυνδεδεμένων

Δεδομένων. Παράλληλα, διασφαλίζουν ότι ακολουθούνται οι Βέλτιστες Πρακτικές για τη Δημοσίευση Διασυνδεδεμένων Δεδομένων (Best Practices for Publishing Linked Data [40]), απαλλάσσουν τους ιδιοκτήτες των δεδομένων από τεχνικές λεπτομέρειες όπως ο τρόπος της ανταλλαγής δεδομένων μεταξύ διαφορετικών συστημάτων και υποστηρίζουν την αντιστοίχιση των URI σε περιγραφές RDF. Επιπλέον, τα περισσότερα από αυτά τα εργαλεία παρέχουν και τη δυνατότητα πρόσβασης στα δεδομένα μέσω SPARQL ερωτημάτων, όπως επίσης και τη δυνατότητα παραγωγής RDF αρχείων με τα δεδομένα που ζητούνται (RDF dumps). Με τη χρήση των εργαλείων αυτών οργανισμοί και ιδιώτες αποκτούν την ικανότητα να παρέχουν τα δεδομένα τους ως Διασυνδεδεμένα Δεδομένα, διευκολύνοντας την ανακάλυψή τους, την πρόσβαση σε αυτά και τη χρήση και ενσωμάτωσή τους σε άλλες εφαρμογές, συμβάλλοντας, έτσι, στην εξάπλωση του Σημασιολογικού Ιστού.

Τα πιο διαδεδομένα εργαλεία δημοσίευσης Διασυνδεδεμένων Δεδομένων είναι τα εξής:

- **D2R Server** [41]: Εργαλείο που επιτρέπει τη δημοσίευση σχεσιακών βάσεων δεδομένων (relational databases) οι οποίες δεν ακολουθούν το πρότυπο RDF σε μορφή Διασυνδεδεμένων Δεδομένων. Ο χρήστης, με τη βοήθεια της δηλωτικής γλώσσας αντιστοίχισης (declarative mapping language) **D2RQ**³⁷, ορίζει τον τρόπο με τον οποίο αντιστοιχίζονται τα περιεχόμενα της βάσης δεδομένων του με το πρότυπο RDF. Έτσι, δεν υπάρχει η ανάγκη για ξεχωριστή αποθήκη RDF δεδομένων, διότι ο D2R Server μετατρέπει άμεσα τα απλά δεδομένα σε Διασυνδεδεμένα μέσω της αντιστοίχισης αυτής. Ο D2R Server επιτρέπει επίσης τη χρήση SPARQL ερωτημάτων στη βάση δεδομένων του.
- **Virtuoso Universal Server**: Ο **Virtuoso**³⁸ είναι μια διεπαφή Διασυνδεδεμένων Δεδομένων για την παροχή RDF δεδομένων και τη χρήση SQL, SPARQL, XQuery και XSLT ερωτημάτων πάνω σε αυτά. Τα RDF δεδομένα μπορούν είτε να είναι αποθηκευμένα στον server είτε να παραχθούν από μη σχεσιακή βάση δεδομένων με βάση κάποια αντιστοίχιση (mapping).
- **Apache Jena Fuseki Server**: Το **Apache Jena**³⁹ είναι μια δομή ανοιχτού κώδικα (open source framework) του Σημασιολογικού Ιστού για τη γλώσσα Java. Παρέχει μια διεπαφή προγραμματισμού εφαρμογών (application programming interface - API) η οποία έχει τη δυνατότητα να δημιουργήσει και να αντλήσει δεδομένα από RDF γράφους. Οι πηγές των δεδομένων των RDF γράφων μπορούν να είναι αρχεία, βάσεις δεδομένων, URLs ή συνδυασμός αυτών. Το Apache Jena υποστηρίζει τη σειριοποίηση των RDF γράφων σε σχεσιακή βάση δεδομένων, καθώς και στις μορφές RDF/XML, Turtle, TriG, Notation3 και JSON-LD. Επιπλέον, υποστηρίζει τη γλώσσα OWL και τη χρήση SPARQL ερωτημάτων και ενημερώσεων στα δεδομένα των RDF γράφων μέσω της μηχανής ερωτημάτων **ARQ**⁴⁰. Με βάση τη δομή Apache Jena αναπτύχθηκε το **Apache Jena Fuseki**⁴¹, μια HTTP διεπαφή για RDF δεδομένα που χρησιμοποιείται ως διακομιστής (server) ή servlet. Η διεπαφή αυτή χρησιμοποιείται στην παρούσα Διπλωματική Εργασία.
- **Blazegraph**: Το **Blazegraph**⁴² είναι μια ανοιχτού κώδικα αποθήκη τριπλετών και βάση δεδομένων γράφων (graph database) σχεδιασμένη για τη διαχείριση μεγάλου όγκου δεδομένων. Υποστηρίζει τη χρήση SPARQL ερωτημάτων. Χρησιμοποιείται από αρκετές γνωστές εταιρίες όπως η **Wikimedia Foundation**⁴³, η **Amazon**⁴⁴ και η **Autodesk**⁴⁵.
- **GraphDB**: Η **Ontotext GraphDB**⁴⁶ είναι μια από τις μεγαλύτερες και γνωστότερες αποθήκες τριπλετών

37 <http://d2rq.org/d2rq-language>

38 <https://virtuoso.openlinksw.com/>

39 <https://jena.apache.org/>

40 <https://jena.apache.org/documentation/query/>

41 <https://jena.apache.org/documentation/fuseki2/>

42 <https://blazegraph.com/>

43 <https://wikimediafoundation.org/>

44 <https://www.aboutamazon.com/>

45 <https://www.autodesk.com/>

46 <https://www.ontotext.com/products/graphdb/>

και βάσεις δεδομένων γράφων. Διαχειρίζεται γράφους γνώσεων και είναι βασισμένη στη δομή **RDF4J**⁴⁷ για την αποθήκευση και ανάλυση RDF δεδομένων. Αποτελείται από τρία βασικά τμήματα: Το **Workbench** για τη διαχείριση της βάσης δεδομένων, το **Engine** που είναι υπεύθυνο για την εξαγωγή συμπερασμάτων, τη βελτιστοποίηση των ερωτημάτων, την αποθήκευση και τη διαχείριση των πρόσθετων και τα **Connectors**, με τη βοήθεια των οποίων η GraphDB συνδέεται και χρησιμοποιεί άλλα λογισμικά ή υπηρεσίες. Η GraphDB ανάμεσα σε πολλά άλλα υποστηρίζει πολλές μορφές RDF σειριοποίησης (όπως RDF/XML, Turtle, N3), τις γλώσσες οντολογιών RDFS, RDFS-plus, OWL2 RL και OWL 2 QL, ερωτήματα σε SPARQL, GraphQL και SeRQL, τη GeoSPARQL, τη χρήση αλγορίθμων εύρεσης των πιο σχετικών οντοτήτων, την πολύπλευρη αναζήτηση (faceted search) και την αναζήτηση με πλήρες κείμενο. Χρησιμοποιείται σε πολλούς σημαντικούς επιστημονικούς τομείς όπως η υγεία, η γενετική, η γεωγραφία, η πολιτιστική κληρονομιά και η ψηφιακή ιστοριογραφία.

2.2.7 Άλλες Τεχνολογίες στο Σημασιολογικό Ιστό

Υπάρχουν πολλές ακόμα τεχνολογίες που χρησιμοποιούνται στο Σημασιολογικό Ιστό. Οι σημαντικότερες από αυτές παρουσιάζονται παρακάτω:

- **Resource Description Framework in Attributes (RDFa)** [42]: Το RDFa είναι ένα πρότυπο που προτείνεται από την W3C και προσθέτει επεκτάσεις που αφορούν τα χαρακτηριστικά στην HTML, την XHTML (Extensible HyperText Markup Language, που είναι XML γλώσσα σήμανσης) και τα αρχεία XML έτσι ώστε να εμπλουτίσει τα έγγραφα ιστού (web documents) με περισσότερα μεταδεδομένα, σύμφωνα με το πρότυπο RDF. Τα μεταδεδομένα αυτά μπορούν να κατανοηθούν και να χρησιμοποιηθούν από υπολογιστές, εφαρμογές, μηχανές αναζήτησης (search engines) και άλλα εργαλεία. Μέσω του RDFa υπάρχει, επίσης, η δυνατότητα εξαγωγής δεδομένων στη μορφή RDF (δηλαδή σε τριπλέτες υποκειμένου-ιδιότητας-αντικειμένου).
- **Microformats (μF)** [43]: Είναι ένα σύνολο ορισμένων HTML κλάσεων που χρησιμοποιούνται ως μεταδεδομένα HTML στοιχείων, περιγράφοντας τον είδος δεδομένων στο οποίο ανήκει το στοιχείο αυτό. Με αυτόν τον τρόπο τα λογισμικά που επεξεργάζονται δεδομένα μπορούν να έχουν μεγαλύτερη αποτελεσματικότητα, καθώς υπάρχουν συγκεκριμένες κλάσεις για τους τύπους δεδομένων και αποφεύγονται οι αυθαίρετοι ορισμοί.
- **Microdata**: Είναι και αυτό πρότυπο για τον εμπλουτισμό των εγγράφων ιστού με μεταδεδομένα. Χρησιμοποιεί ζεύγη ονόματος-τιμής από γνωστά λεξιλόγια ή λεξιλόγια φτιαγμένα για τον σκοπό της εκάστοτε εργασίας (προτιμώνται πάντα τα γνωστά λεξιλόγια για λόγους επαναχρησιμοποίησης περιεχομένου) και είναι πιο απλός τρόπος πρόσθεσης μεταδεδομένων σε HTML στοιχεία σε σύγκριση με το RDFa και το Microformats.
- **Shapes Constraint Language (SHACL)** [44]: Είναι γλώσσα περιγραφής RDF γράφων και αποτελεί W3C πρότυπο για την επικύρωσή τους. Τα μοντέλα SHACL μπορούν να εκφραστούν σε όλες τις μορφές RDF σειριοποίησης και θέτουν περιορισμούς στη δομή, το περιεχόμενο και τη σημασιολογία των RDF γράφων. Ορίζουν ιδιότητες όπως η καρδιναιότητα μιας ιδιότητας, ο τύπος και το εύρος των τιμών που παίρνει μια ιδιότητα κ.λ.π.. Για την επικύρωση ενός RDF γράφου, η SHACL παίρνει ως είσοδο το γράφο αυτό (που ονομάζεται data graph) και ένα SHACL μοντέλο (που ονομάζεται shapes graph) και παράγει μία έκθεση επικύρωσης (validation report), στην οποία περιλαμβάνεται ο βαθμός της σοβαρότητας κάθε «λάθους», όπως επίσης και προτάσεις για την διόρθωση των δεδομένων αν τα λάθη είναι σημαντικά. Μάλιστα, κάποιοι SHACL επεξεργαστές έχουν τη δυνατότητα για αυτόματη διόρθωση των λαθών με βάση τις προτάσεις αυτές. Επιπλέον, μέσω του SHACL Rules χαρακτηριστικού της, το οποίο είναι εμπνευσμένο από αντίστοιχο χαρακτηριστικό SPIN Rules της SPIN, η SHACL έχει επίσης δυνατότητες συλλογισμού, αντιστοίχισης και

47 <https://rdf4j.org/>

μετατροπής δεδομένων.

- Επεξεργαστές Οντολογιών (Ontology Editors): Είναι λογισμικά για τη δημιουργία και επεξεργασία οντολογιών, με διαφορετικά επιπλέον χαρακτηριστικά και εργαλεία ανάλογα με τον σκοπό και τις δυνατότητες που προσφέρει ο καθένας. Οι πιο γνωστοί από αυτούς είναι ο **Topbraid Composer**⁴⁸, ο **Protégé**⁴⁹ [45] (ο οποίος είναι ο πιο διαδεδομένος επεξεργαστής οντολογιών παγκοσμίως, και χρησιμοποιήθηκε και από την παρούσα Διπλωματική Εργασία) και το **NeOn Toolkit**⁵⁰ [46] (για πιο μεγάλες εφαρμογές).
- Σημασιολογικοί Επαγωγείς (Semantic Reasoners): Αποτελούν λογισμικά τα οποία έχουν στόχο να καταλήξουν σε λογικά συμπεράσματα μέσα από την ανάλυση αξιωμάτων. Παράγουν νέα γνώση που δεν αναφέρεται ρητά στα δεδομένα της οντολογίας. Οι σημασιολογικοί επαγωγείς έχουν, μεταξύ άλλων, ως αντικείμενο να ελέγχουν αν η οντολογία είναι συνεπής ή έρχεται σε αντιφάσεις οι οποίες πρέπει να διορθωθούν, να εξετάζουν το κατά πόσο είναι γενικότερη μια έννοια που περιγράφεται στην οντολογία σε σχέση με κάποια άλλη και να εκτελούν εργασίες όπως η ανάκτηση όλων τα ατόμων μιας κλάσης ή όλων των κλάσεων στις οποίες ανήκει ένα άτομο. Συχνά οι επεξεργαστές οντολογιών εμπεριέχουν σημασιολογικούς επαγωγείς για τον έλεγχο της οντολογίας που αναπτύσσει ο χρήστης. Ο επεξεργαστής Protégé χρησιμοποιεί τον επαγωγέα **Hermit**⁵¹ [47], ενώ οι επαγωγείς **Pellet**⁵² [48] και **FaCT++**⁵³ [49] είναι διαθέσιμοι ως πρόσθετα (plug-ins).

2.3 Το Πρότυπο ELI

Η θέσπιση του προτύπου ELI (European Legislation Identifier - Αναγνωριστικό Ευρωπαϊκής Νομοθεσίας) [50] αποτελεί μια πρωτοβουλία της Ευρωπαϊκής Ένωσης με στόχο τη διευκόλυνση της συνεργασίας των κρατών-μελών της μέσω του εκσυγχρονισμού της ταυτοποίησης, της πρόσβασης, της ανταλλαγής και της επαναχρησιμοποίησης νομικών πληροφοριών μεταξύ των δημοσίων αρχών, των επαγγελματιών στο νομικό τομέα, των ακαδημαϊκών και των απλών πολιτών. Η σύλληψη της ιδέας για την ανάπτυξη του ELI προήλθε από την αναγνώριση του Παγκοσμίου Ιστού ως ένα εργαλείο που δίνει νέες δυνατότητες στην πρόσβαση, τον διαμοιρασμό και τον εμπλουτισμό νομικών πληροφοριών. Η πρωτοβουλία αυτή στηρίζεται από το Ευρωπαϊκό Φόρουμ Επίσημων Εφημερίδων (**European Forum of Official Gazettes**⁵⁴) και την Υπηρεσία Εκδόσεων της Ευρωπαϊκής Ένωσης (**Publications Office of the European Union**⁵⁵), ενώ η επιτροπή ELI Task Force (ELI TF) είναι υπεύθυνη για την ανάπτυξη και την προώθηση της χρήσης του προτύπου ELI, με στόχο την ευκολότερη και ευρύτερη εφαρμογή του από τις χώρες της Ευρωπαϊκής Ένωσης.

2.3.1 Τεχνικά Στοιχεία

Το ELI χρησιμοποιείται για την ταυτοποίηση της νομοθεσίας μέσω ενός μοναδικού ταυτοποιητή, κατανοητού από μηχανές και ανθρώπους και συμβατού με τα νεότερα τεχνολογικά πρότυπα στο Σημασιολογικό Ιστό και για την περιγραφή της νομοθεσίας μέσω ενός συνόλου κατανοητών από μηχανές στοιχείων μεταδεδομένων, συμβατών με μια προτεινόμενη οντολογία. Είναι ένα λεξιλόγιο με βάση το οποίο

48 <https://archive.topquadrant.com/topbraid-composer-install/>

49 <https://protege.stanford.edu/>

50 http://neon-toolkit.org/wiki/Main_Page.html

51 <http://www.hermit-reasoner.com/>

52 <https://github.com/stardog-union/pellet>

53 <http://owl.cs.manchester.ac.uk/tools/fact/>

54 <https://op.europa.eu/en/web/forum>

55 <https://op.europa.eu/en/home>

αναπαρίστανται μεταδεδομένα που αφορούν την νομοθεσία της Ευρωπαϊκής Ένωσης και τις εθνικές νομοθεσίες των κρατών-μελών της, χωρίς αυτό να αποκλείει τη χρήση του από άλλους οργανισμούς. Προσφέρει έναν τυποποιημένο τρόπο ταυτοποίησης και περιγραφής του πλαισίου και του περιεχομένου της νομοθεσίας δίνοντας χρήσιμες πληροφορίες όπως το πεδίο εφαρμογής της, ο σκοπός της και η σχέση της με άλλες νομοθεσίες. Η ανάπτυξη του ELI αποτελεί το έναυσμα για την ανάπτυξη σημασιολογικών γράφων γνώσης στο νομικό πεδίο.

Οι τρεις Άξονες (Pillars) του ELI, όπως είναι ορισμένοι από την ELI TF, είναι οι εξής:

- **Pillar I: Ταυτοποιητής ELI (ELI Identifier)**

Το ELI χρησιμοποιεί URI πρότυπα τα οποία έχουν νομική σημασιολογία και, ταυτόχρονα, είναι φιλικά προς το χρήστη. Τα πρότυπα αυτά είναι μια σειρά από συστατικά στοιχεία, όσα χρειάζεται από τα οποία κάθε κράτος μπορεί να χρησιμοποιήσει με σκοπό να δημιουργήσει τα δικά του URI, λαμβάνοντας υπόψιν τις δικές του ιδιαιτερότητες και ανάγκες. Τα URI που δημιουργούνται κατ' αυτόν τον τρόπο λειτουργούν ως παγκόσμιοι ταυτοποιητές της νομοθεσίας. Παράδειγμα ενός τέτοιου URI που ταυτοποιεί ένα φύλλο της Επίσημης Εφημερίδας (Official Journal - oj) της Ευρωπαϊκής Ένωσης είναι το εξής:

http://example.eu/eli/{typedoc}/{year}/{natural_number}/oj

Τα συστατικά στοιχεία των URI του προτύπου ELI είναι τα εξής:

Όνομα	Σχόλια
eli	
Jurisdiction	Προσδιορισμός της περιοχής που αφορά η πηγή νομοθεσίας. Για τις χώρες χρησιμοποιούνται οι διψήφιοι κωδικοί χωρών όπως ορίζονται στο DCTERMS.ISO3166 (π.χ. gr), ενώ για οργανισμούς το επίσημο αναγνωριστικό τους (π.χ. EU).
Agent	Διοικητική ιεραρχική δομή (π.χ. Κοινοβούλιο).
Subagent	Διοικητική δομή κατώτερη στην ιεραρχία (π.χ. Υπουργείο).
Year	Έτος στη μορφή EEEE. Μπορεί, ανάλογα με τις ανάγκες κάθε κράτους, να αφορά ημερομηνία υπογραφής, έκδοσης κ.λ.π..
Month	Μήνας σε μορφή MM.
Day	Ημέρα σε μορφή HH.
Type	Τύπος της πηγής νομοθεσίας (π.χ. Νόμος, Προεδρικό Διάγραμμα, Εγκύκλιος, Απόφαση κ.λ.π.). Ποικίλει ανάλογα με τις ανάγκες κάθε κράτους.
Subtype	Υποκατηγορία πηγής νομοθεσίας ανάλογα με τις ανάγκες κάθε κράτους.
Domain	Η θεματολογία της πηγής νομοθεσίας αν αυτή ταξινομείται με βάση το θέμα της (π.χ. με κωδικούς)
Natural Identifier	Αριθμός που ξεχωρίζει την πηγή νομοθεσίας από τις υπόλοιπες πηγές νομοθεσίας ίδιου τύπου που υπογράφηκαν ή εκδόθηκαν την ίδια μέρα.
Level 1	Υποκατηγορία της πηγής νομοθεσίας (π.χ. Κεφάλαιο).
Level 2	Υποκατηγορία της προηγούμενης υποκατηγορίας (π.χ. Άρθρο).
Level 3	Υποκατηγορία της προηγούμενης υποκατηγορίας (π.χ. Παράγραφος).
Level n	Υποκατηγορία της προηγούμενης υποκατηγορίας (συνεχίζεται n φορές...).
Point in Time	Ημερομηνία κατά την οποία είναι έγκυρη η πηγή νομοθεσίας σε μορφή EEEEMMHH.
Version	Ξεχωρίζει την αρχική μορφή τη πηγής νομοθεσίας από άλλες μορφές της (π.χ. σύνοψη).
Language	Η γλώσσα στην οποία είναι γραμμένες οι διάφορες εκδόσεις της ίδιας πηγής νομοθεσίας. Χρησιμοποιούνται οι τριψήφιοι κωδικοί χωρών όπως ορίζονται στο DCTERMS.ISO3166 (π.χ. grc).

Πίνακας 1: Συστατικά Στοιχεία των URI του προτύπου ELI

Η ELI TF στον οδηγό της, «ELI technical guide [51]», προτείνει, μεταξύ άλλων, τα URIs που δημιουργούνται από τους εκάστοτε προγραμματιστές να είναι εύκολα κατανοητά, επεξεργάσιμα και απομνημονεύσιμα από τους χρήστες, να περιλαμβάνουν όσα στοιχεία χρειάζεται ώστε κάθε πηγή νομοθεσίας να ορίζεται ξεχωριστά χωρίς όμως να υπάρχουν περιττά στοιχεία, και τα στοιχεία που χρησιμοποιούν να παραμένουν σταθερά στη διάρκεια του χρόνου. Επιπλέον, στον ίδιο οδηγό, ορίζει τα «Ελλιπή ELIs» (Incomplete/Partial ELIs) και προτείνει την υποστήριξή τους από τις ELI οντολογίες που αναπτύσσονται. Με βάση το προηγούμενο παράδειγμα, ένα ελλιπές ELI θα ήταν της μορφής <http://example.eu/eli/{typedoc}/{year}> και θα επέστρεφε όλα τα αρχεία του τύπου που ορίζει το κομμάτι typedoc και δημοσιεύτηκαν τον χρόνο που ορίζει το κομμάτι year. Συμπεραίνουμε ότι τα Ελλιπή ELIs είναι, ουσιαστικά, τα αρχικά τμήματα των URIs της οντολογίας και, ενώ δεν οδηγούν σε συγκεκριμένα αρχεία νομοθεσίας, μπορούν να λειτουργήσουν ως ένα εργαλείο αναζήτησης που θα διευκολύνει τον εκάστοτε χρήστη στην έρευνά του.

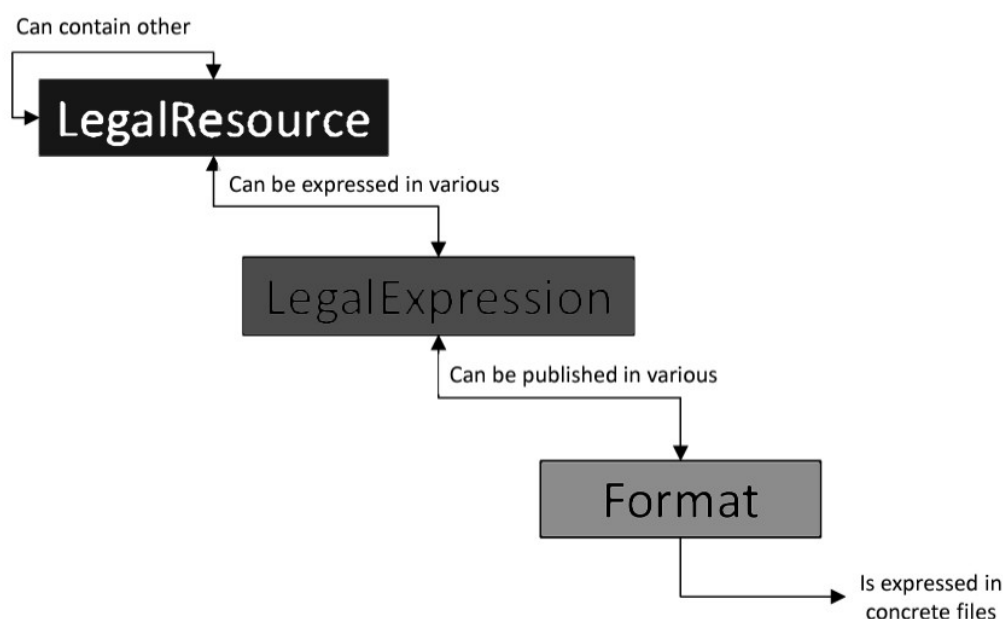
- Pillar II: ELI Μεταδεδομένα και Οντολογία (ELI Metadata and Ontology)

Η ELI οντολογία ορίζει ένα κοινό μοντέλο δεδομένων για την ανταλλαγή νομικών μεταδεδομένων στο διαδίκτυο. Η περιγραφή της νομοθεσίας στο πρότυπο ELI ακολουθεί τις αρχές του **FRBR**⁵⁶ (Functional Requirements for Bibliographic Records), το οποίο είναι ένα μοντέλο οντοτήτων-συσχετίσεων (entity-relationship model) που αφορά διαδικτυακούς καταλόγους βιβλιοθηκών και βάσεις δεδομένων βιβλιογραφίας. Επιπλέον, οι ιδιότητες της ELI οντολογίας αντιστοιχίζονται σε μεταδεδομένα του λεξιλογίου Dublin Core, το οποίο αποτελεί ένα σύνολο αντικειμένων μεταδεδομένων που περιγράφουν φυσικές ή διαδικτυακές πηγές.

Η οντολογία του προτύπου ELI είναι εύκαμπτη και αρκετά γενική, έτσι ώστε να μπορεί να υποστηρίξει τις ιδιαιτερότητες κάθε νομοθεσίας που τη χρησιμοποιεί χωρίς να θυσιάζεται η διαλειτουργικότητα των νομικών μεταδεδομένων. Το πρότυπο ELI επιτρέπει στον προγραμματιστή να προσθέσει στην ELI οντολογία που αναπτύσσει όσες νέες κλάσεις ή ιδιότητες χρειάζεται για το έργο του, είτε πρόκειται για κλάσεις/ιδιότητες που προέρχονται από άλλες οντολογίες, είτε για κλάσεις/ιδιότητες που ο ίδιος δημιούργησε. Η ELI οντολογία είναι, επιπλέον, απλή στη χρήση και καλά τεκμηριωμένη, έτσι ώστε να διευκολύνεται η εφαρμογή της από τα κράτη-μέλη της Ευρωπαϊκής ένωσης, ενώ, δεδομένου ότι βασίζεται σε ήδη υπάρχουσες τεχνολογίες (RDF, OWL) και λεξιλόγια (FRBR, SKOS, Dublin Core), είναι επίσης ανοιχτή.

Η κεντρική μορφή της οντολογίας ELI συνοψίζεται με τις εξής 3 κλάσεις:

56 <https://www.loc.gov/catdir/cpsd/frbrengr.pdf>



Εικόνα 7: Η βασική μορφή του πρότυπου ELI

Η κλάση «**LegalResource**» βασίζεται στην αντίστοιχη κλάση «**Work**» του FRBR και περιγράφει τα νομοθετικά δημιουργήματα σαν έννοιες, όπως για παράδειγμα ο Νόμος 3918/2011, χωρίς να δίνει στοιχεία για τη μορφή τους, το περιεχόμενό τους ή τη γλώσσα στην οποία είναι γραμμένα. Η ίδια κλάση μπορεί να χρησιμοποιηθεί ως υποκλάση στον εαυτό της για να περιγράψει τμήματα του νομοθετικού δημιουργήματος, όπως κεφάλαια, άρθρα κ.λ.π.. Η κλάση «**LegalExpression**» βασίζεται στην αντίστοιχη κλάση «**Expression**» του FRBR και περιγράφει «εκφράσεις» ενός νομοθετικού δημιουργήματος, όπως για παράδειγμα το περιεχόμενο του Νόμου 3918/2011 στα αγγλικά ή στα ελληνικά. Κάθε **LegalExpression** οφείλει να συνδέεται με το **LegalResource** που περιγράφει μέσω της ιδιότητας `eli:realizes`. Τέλος, η κλάση «**Format**» βασίζεται στην αντίστοιχη κλάση «**Manifestation**» του FRBR και είναι η φυσική ενσωμάτωση, είτε σε χαρτί είτε ηλεκτρονική μορφή, μιας έκφρασης νομοθετικού δημιουργήματος, όπως, για παράδειγμα, ο Νόμος 3918/2011 στα αγγλικά σε μορφή PDF. Κάθε **Format** πρέπει να συνδέεται με το **LegalExpression** που ενσωματώνει μέσω της ιδιότητας `eli:embodies`. Η πλήρης λίστα των κλάσεων, των ιδιοτήτων και των χαρακτηριστικών που αποτελούν την ELI οντολογία, καθώς και η εξήγησή τους περιγράφεται αναλυτικά στον **Πίνακα Μεταδεδομένων της Οντολογίας ELI**⁵⁷.

• Pillar III: Ενσωμάτωση ELI μεταδεδομένων σε νομικές ιστοσελίδες (Integration of ELI metadata into legislative websites)

Τα μεταδεδομένα ELI πρέπει να κωδικοποιούνται είτε με βάση το πρότυπο RDFa, είτε σε μορφή JSON-LD. Οι προγραμματιστές είναι ελεύθεροι, αν το επιθυμούν, να χρησιμοποιήσουν επιπλέον μορφές σειριοποίησης στις εφαρμογές τους. Ανάμεσα στις δυνατότητες που έχουν περιλαμβάνεται η διάθεση των ELI μεταδεδομένων σε μορφή RDF αρχείων και η δημιουργία RDF βάσης δεδομένων η οποία θα είναι προσβάσιμη στο κοινό μέσω SPARQL ερωτημάτων.

Για την εφαρμογή του προτύπου ELI, το ελάχιστο που οι προγραμματιστές οφείλουν να κάνουν είναι να δώσουν URIs στα νομοθετικά δημιουργήματα που αφορούν την έργο που έχουν αναλάβει, χωρίς να είναι απαραίτητο το έργο αυτό να αφορά όλη την υπαρκτή νομοθεσία. Τα URIs αυτά θα είναι πολλά για κάθε νομοθετικό δημιούργημα, καθώς θα υπάρχει ένα URI για το **LegalResource** του, ένα για κάθε **LegalExpression** του και ένα για κάθε **Format** του. Επιπλέον, είναι θεμιτό να δοθούν ξεχωριστά URIs για κάθε μέρος (όπως άρθρο, κεφάλαιο κ.λ.π.) του νομοθετικού δημιουργήματος.

⁵⁷ https://op.europa.eu/documents/3938058/11669184/eli_ontology.xlsx

2.3.2 Επεκτάσεις και Εργαλεία της Οντολογίας ELI

- **ELI Ontology for Draft Legislation (ELI-DL):** Η **ELI-DL**⁵⁸ αποτελεί επέκταση της ELI οντολογίας που αφορά προσχέδια νομοθεσιών, με σκοπό τη διάδοση δομημένων δεδομένων για νομοθεσίες που βρίσκονται υπό επεξεργασία και δεν έχουν ακόμα δημοσιευτεί. Περιλαμβάνει τρεις κύριες έννοιες: το **LegislativeProject**, που περιγράφει τη σχεδιαζόμενη νομοθεσία από τη στιγμή που προτάθηκε μέχρι την τελική της δημοσίευση, το **LegalActivity**, το οποίο περιγράφει τις διαδικασίες από τις οποίες «περνά» η σχεδιαζόμενη νομοθεσία και το **LegislativeProjectWork**, το οποίο αφορά τα αρχεία που δημιουργούνται κατά τη διάρκεια του σχεδιασμού της νομοθεσίας. Στόχος της ELI-DL είναι η ευκολότερη και πιο έγκαιρη ανταλλαγή δεδομένων μεταξύ νομικών συστημάτων, η δυνατότητα πρόβλεψης των επιπτώσεων της σχεδιαζόμενης νομοθεσίας σε υπάρχουσες νομοθεσίες, η παρακολούθηση της διαδικασίας σχεδιασμού της νέας νομοθεσίας, η έγκαιρη ενημέρωση πάνω σε αυτή και η αύξηση της διαφάνειας προς τους πολίτες.
- **ELI Annotation Tool:** Το **ELI Annotation Tool**⁵⁹ είναι μια εφαρμογή που επιτρέπει στους εκδότες νομικής πληροφορίας να ταυτοποιήσουν και να περιγράψουν τις πηγές νομοθεσίας τους σύμφωνα με το πρότυπο ELI, καθώς επίσης και να διαθέσουν τα δεδομένα αυτά στο διαδίκτυο ενσωματωμένα σε HTML σελίδες. Αποτελείται από δύο μέρη, ένα για τη διαχείριση του μοντέλου δεδομένων και των ελεγχόμενων λεξιλογίων, που χρησιμοποιείται για τη προσαρμογή της εφαρμογής και των URI στις ανάγκες του εκάστοτε έργου, και ένα για την επεξεργασία και τη δημοσίευση των μεταδεδομένων των νομικών πηγών, που χρησιμοποιείται για την έκδοση HTML σελίδων με μεταδεδομένα στις μορφές RDFa και JSON-LD, αλλά και σε μορφή κατάλληλη για ανάγνωση από ανθρώπους.
- **ELI Validator:** Το **ELI Validator**⁶⁰ είναι μια διαδικτυακή υπηρεσία με την οποία ο προγραμματιστής μπορεί να ελέγξει κατά πόσο η RDFa σήμανση μιας HTML ιστοσελίδας ακολουθεί το πρότυπο ELI. Παράγει μια αναφορά αξιολόγησης που περιλαμβάνει Infos (μηνύματα με στόχο να βελτιώσουν την ποιότητα των ELI μεταδεδομένων), Warnings (μηνύματα για μικρές παραβάσεις οι οποίες δε θα εμποδίσουν την κατανόηση των μεταδεδομένων από τον χρήστη) και Violations (μηνύματα για σοβαρότερες παραβάσεις που θα εμποδίσουν τη σωστή κατανόηση των μεταδεδομένων από το χρήστη).
- **ELI/XML:** Αποτελεί μια ανεπίσημη κωδικοποίηση με στόχο να βοηθηθεί η έκφραση ELI μεταδεδομένων σε αρχεία XML. Διαθέτει XML μετασχηματισμούς για την παραγωγή ELI μεταδεδομένων στη μορφή RDF/XML. Έχει στόχο να διευκολύνει την ενσωμάτωση του προτύπου ELI σε ροές εργασιών που βασίζονται στη χρήση XML αρχείων.

58 <https://joinup.ec.europa.eu/collection/eli-european-legislation-identifier/solution/eli-ontology-draft-legislation-eli-dl/about>

59 <https://joinup.ec.europa.eu/collection/eli-european-legislation-identifier/solution/eli-annotation-tool/distribution/eli-annotation-tool>

60 <https://webgate.ec.europa.eu/eli-validator/validate>

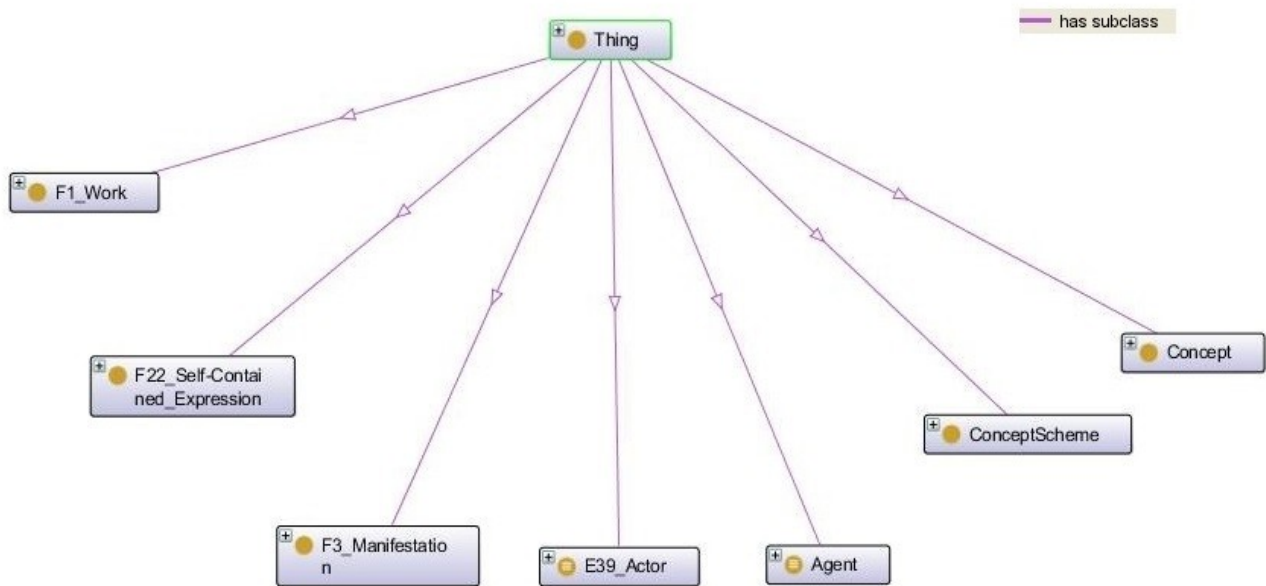
3

Αναπαράσταση Ελληνικής Νομοθεσίας

3.1 Μοντέλο Διασυνδεδεμένων Δεδομένων Ελληνικής Νομοθεσίας

Ο πρώτος στόχος της παρούσας Διπλωματικής Εργασίας είναι η ανάπτυξη ενός μοντέλου Διασυνδεδεμένων Δεδομένων για την ελληνική νομοθεσία με βάση την **οντολογία ELI**⁶¹ της Ευρωπαϊκής Ένωσης, την οποία παρέχει σε OWL μορφή η Υπηρεσία Εκδόσεων της Ευρωπαϊκής Ένωσης. Για τη χρήση του προτύπου ELI στη δημοσίευση της ελληνικής νομοθεσίας με μορφή Διασυνδεδεμένων Δεδομένων κρίνεται απαραίτητο το ELI να τροποποιηθεί με τον κατάλληλο τρόπο, έτσι ώστε να ανταποκρίνεται στην σύγχρονη ελληνική πραγματικότητα. Για το λόγο αυτό, και με τη χρήση του επεξεργαστή οντολογιών Protégé, έγιναν ορισμένες προσθήκες απαραίτητων κλάσεων και σχέσεων που δεν υπήρχαν στο πρότυπο ELI. Με αυτόν τον τρόπο, δημιουργήσαμε το ELIgr, το οποίο αποτελεί επέκταση του προτύπου ELI. Το ELIgr, για να μπορεί να είναι συμβατό με το ELI, διατηρεί όλες τις κλάσεις και τις ιδιότητες του ELI. Οι βασικές κλάσεις του προτύπου ELI είναι οι εξής (όλα τα σχήματα που ακολουθούν στην παρούσα Ενότητα δημιουργήθηκαν με τον επεξεργαστή οντολογιών Protégé):

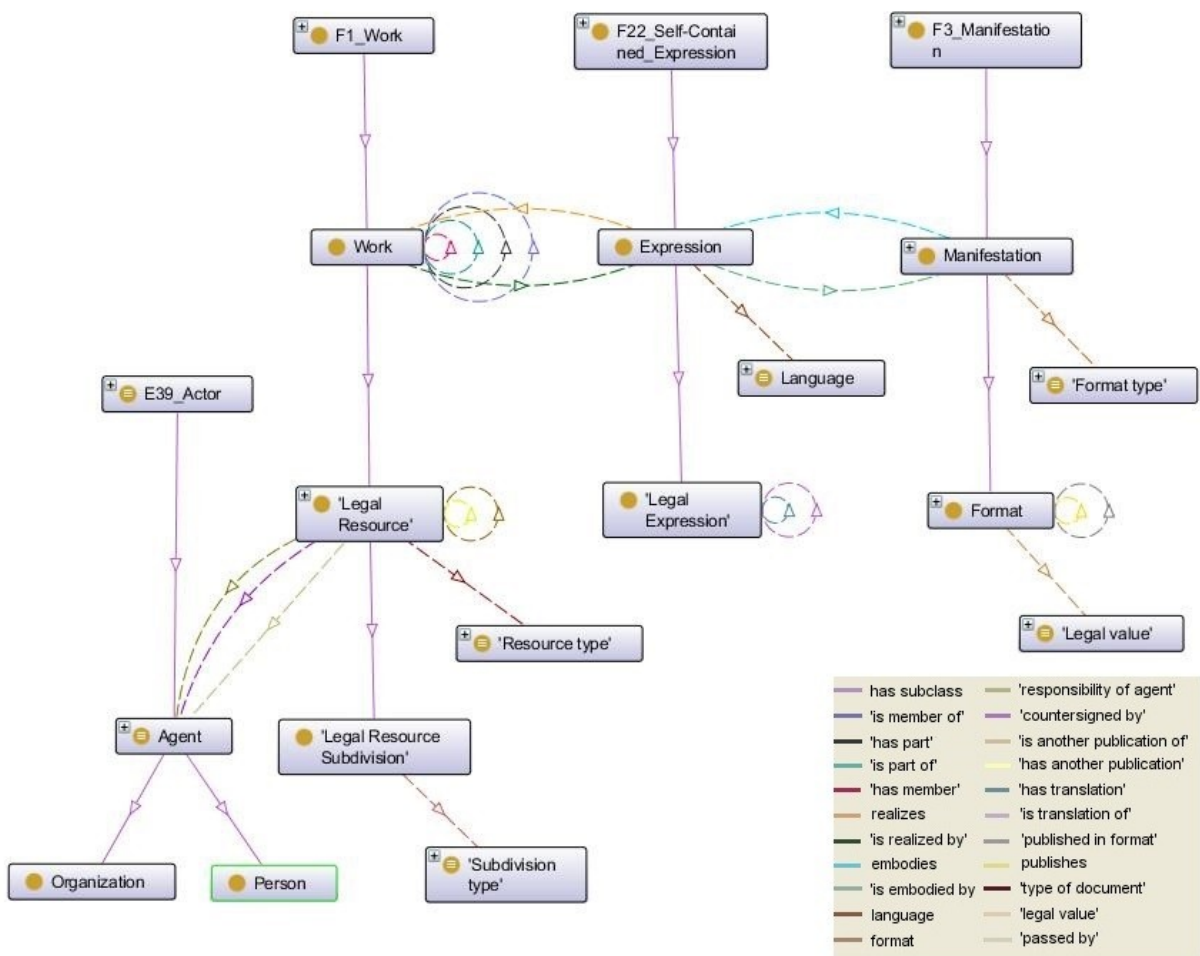
61 <https://op.europa.eu/documents/3938058/11669184/eli.owl/>



Εικόνα 8: Οι βασικές κλάσεις της οντολογίας ELI

Από αυτές τις κλάσεις μπορούμε να θεωρήσουμε ως σημαντικότερες τις F1_Work, F22_Self-Contained_Expression και F3_Manifestation. Οι τρεις αυτές κλάσεις προέρχονται από το λεξιλόγιο FRBR, ενώ οι κλάσεις Concept και ConceptScheme από το λεξιλόγιο SKOS. Η κλάση E39_Actor προέρχεται επίσης από το FRBR και είναι ισοδύναμη της κλάσης Agent, η οποία είναι η μόνη από τις βασικές κλάσεις που είναι εγγενής της οντολογίας ELI και δεν προέρχεται από άλλα λεξιλόγια.

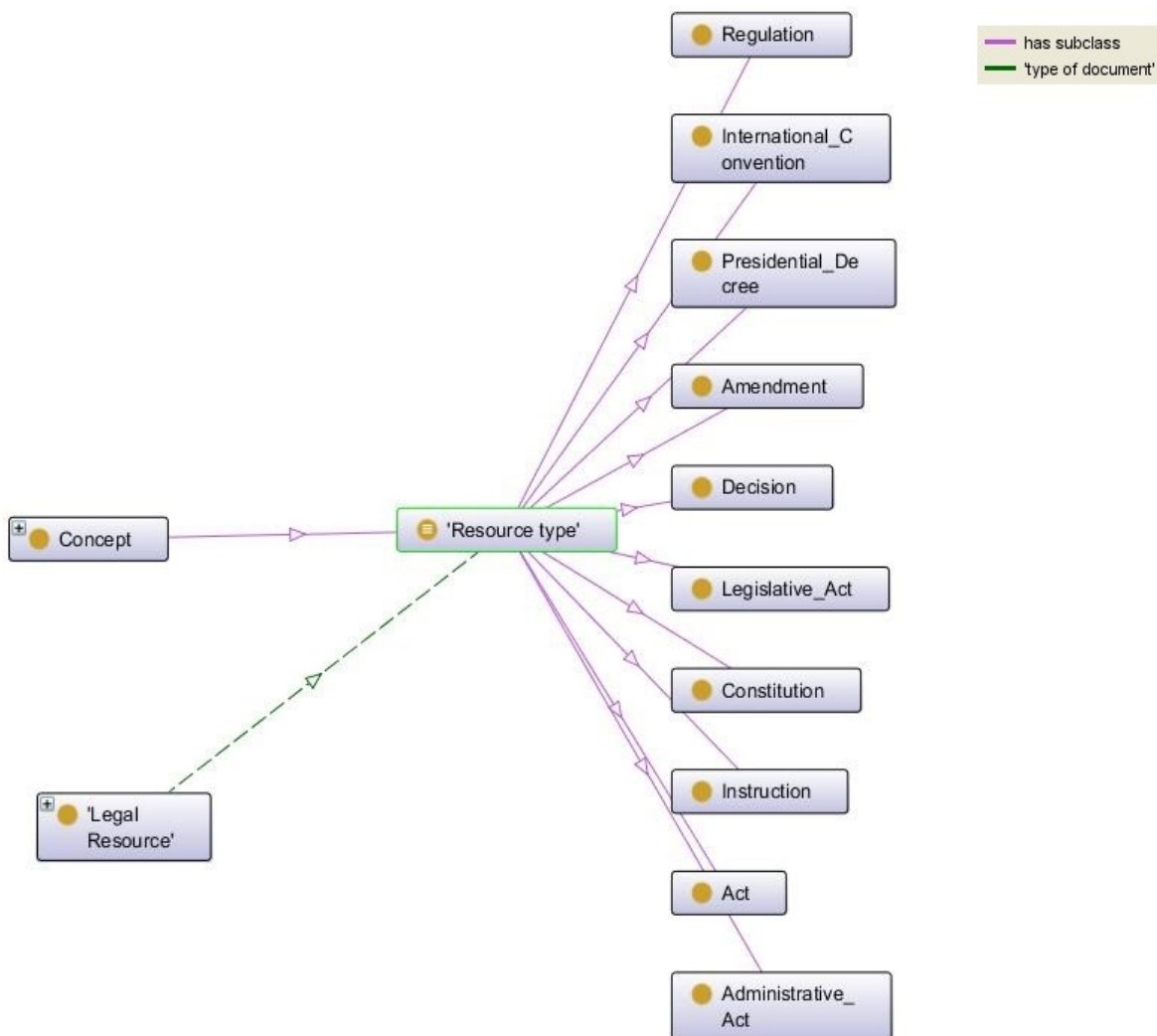
Ένας γράφος που περιγράφει αναλυτικότερα τις σημαντικότερες κλάσεις της οντολογίας ELI και τις μεταξύ τους σχέσεις είναι ο παρακάτω:



Εικόνα 9: Οι κλάσεις της οντολογίας ELI και οι μεταξύ τους σχέσεις

Όπως αναφέραμε, η επέκταση ELIgr περιέχει όλες τις παραπάνω κλάσεις και σχέσεις του προτύπου ELI, έτσι ώστε να είναι συμβατή με αυτό. Οι επιπλέον προσθήκες που κάναμε για την προσαρμογή του ELI στα ελληνικά δεδομένα είναι οι παρακάτω:

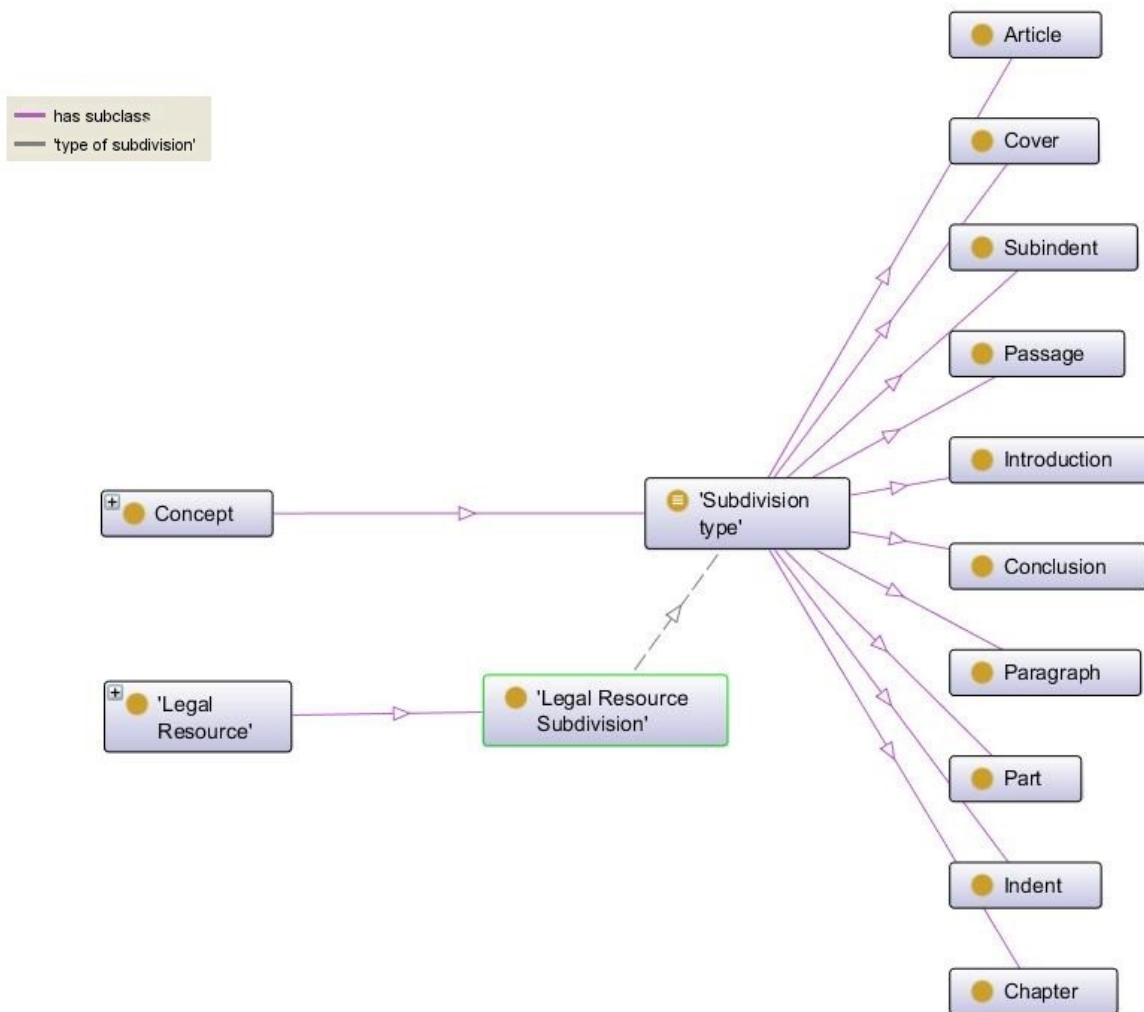
- Στο Concept → Resource type προστέθηκαν ως ασυμπτωτικές μεταξύ τους υποκλάσεις τα Act (Τυπικός Νόμος), Administrative_Act (Διοικητική Πράξη), Amendment (Τροπολογία), Constitution (Σύνταγμα), Decision (Απόφαση), Instruction (Οδηγία/Εγκύκλιος), International_Convention (Διεθνής Σύμβαση), Legislative_Act (Νομοθετική Πράξη), Presidential_Decree (Προεδρικό Διάταγμα), και Regulation (Κανονισμός). Οι κατηγορίες αυτές είναι τα είδη νομικών πράξεων στο ελληνικό δίκαιο σύμφωνα με την [ιστοσελίδα του Οδηγού Νομικών Υπηρεσιών](#)⁶² στην Ελλάδα.



Εικόνα 10: Τα Είδη των Νομικών Πράξεων στο ELIgr

62 <http://www.greekjustice.gr/>

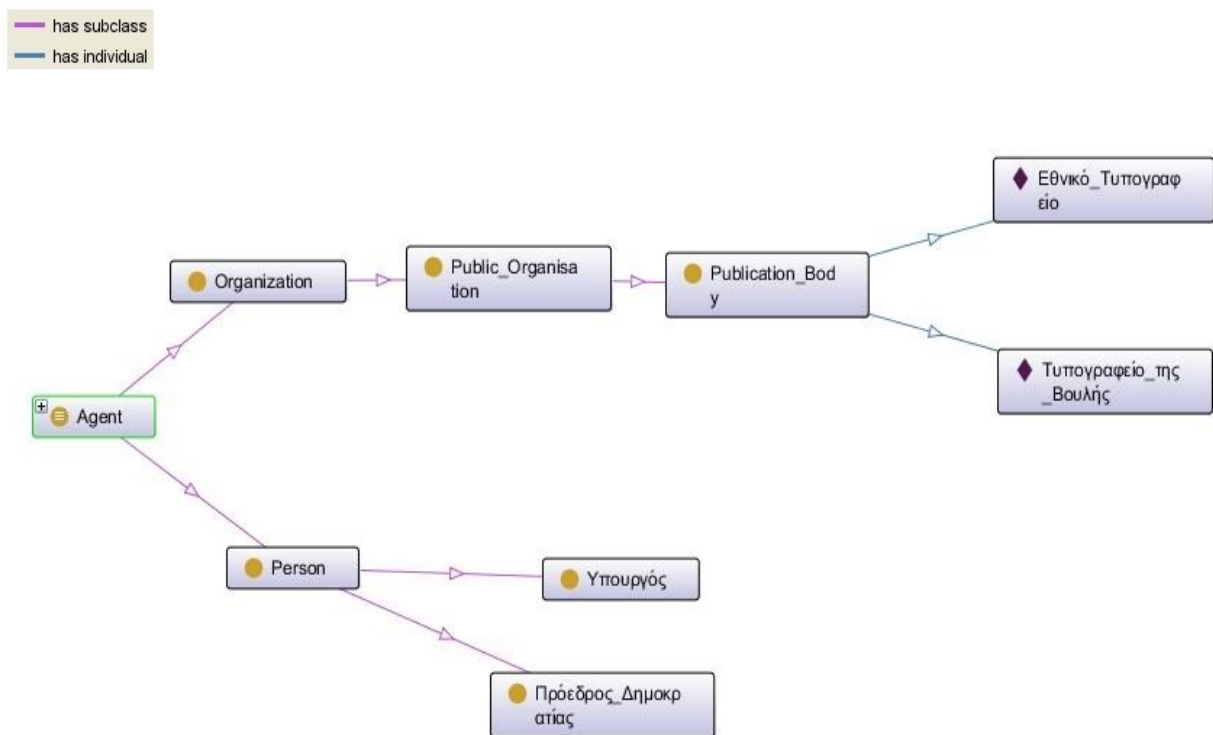
- Στο Concept → Subdivision type προστέθηκαν ως ασυμπτωτικές μεταξύ τους υποκλάσεις τα Article (Άρθρο), Chapter (Κεφάλαιο), Conclusion (Επίλογος), Cover (Εξώφυλλο), Indent (Περίπτωση), Introduction (Εισαγωγή), Paragraph (Παράγραφος), Part (Μέρος), Passage (Εδάφιο) και Subindent (Υποπερίπτωση) σύμφωνα με την **Οντολογία Κωδικοποιήσεων που προτείνει το Υπουργείο Εσωτερικών**⁶³ και με βάση τις ανάγκες που προέκυψαν κατά τη δημιουργία των ελληνικών νομικών σημασιολογικών εγγράφων σε RDF μορφή.



Εικόνα 11: Τα Είδη των Υποδιαίρέσεων Νομικών Πράξεων στο ELIgr

63 http://www.opengov.gr/types/wp-content/uploads/downloads/2016/07/2016_06_29ontology.pdf

- Στο F1_Work → Work → Legal Resource προστέθηκε το ΦΕΚ ως υποκλάση. Στη συνέχεια, προστέθηκε το Public_Organisation και το subclass του, Publication_Body, ως subclass στο Agent → Organisation. Έπειτα, προστέθηκε το Εθνικό_Τυπογραφείο και το Τυπογραφείο_της_Βουλής ως άτομο του Publication_Body (Agent → Organization → Public_Organisation → Publication_Body). Προστέθηκε και το site του Εθνικού Τυπογραφείου (<http://www.et.gr>) ως "homepage". Για να γίνει αυτό, δημιουργήθηκε νέο data property με όνομα "homepage", πεδίο ορισμού Organisation και πεδίο τιμών string. Οι προσθήκες αυτές έγιναν σύμφωνα με την Οντολογία Κωδικοποιήσεων που προτείνει το Υπουργείο Εσωτερικών. Εφόσον το Εθνικό_Τυπογραφείο είναι άτομο του Publication_Body (υποκλάση του Agent) και το ΦΕΚ είναι υποκλάση του Legal Resource, συνδέονται μεταξύ τους με την ιδιότητα publisher agent.
- Στο Agent → Person προστέθηκαν ως υποκλάσεις τα Πρόεδρος_Δημοκρατίας και Υπουργός



Εικόνα 12: Οι προσθήκες στην κλάση Agent στο ELIgr

- Στο F3_Manifestation → Manifestation → Format προστέθηκε σαν υποκλάση η Εφημερίδα της Κυβέρνησης (με όνομα officialGazette), σύμφωνα, πάλι, με την Οντολογία Κωδικοποιήσεων που προτείνει το Υπουργείο Εσωτερικών.
- Στην ιδιότητα changes → amends προστέθηκαν οι υποϊδιότητες inserts (προσθέτει), joins (συνενώνει), renumbers (επαναριθμεί), splits (διαχωρίζει), substitutes (αντικαθιστά) σύμφωνα με την ίδια Οντολογία Κωδικοποιήσεων. Όλες αυτές οι υποϊδιότητες είναι ασυμπτωτικές μεταξύ τους και έχουν πεδία ορισμού και τιμών τα Legal Resource και Legal Expression. Στο property changed by → amended by προστέθηκαν οι αντίστροφες ιδιότητες των παραπάνω (is_inserted_by, is_joined_by, is_renumbered_by, is_split_by, is_substituted_by).
- Οι ιδιότητες corrects, cites, changes, amends, repeals, commences, inserts, joins, renumbers, splits, substitutes, consolidates, realizes, embodies, basis_for, has part, has member, publishes και όλες οι αντίστροφές τους χαρακτηρίστηκαν ως ασύμμετρες, δεδομένου ότι είναι αδύνατον να αντιστραφούν τα

υποκείμενα και τα αντικείμενα τα οποία συνδέουν κάθε φορά. Για παράδειγμα, είναι αδύνατο μια νομοθεσία Β να αλλάζει μια νομοθεσία Α, και ταυτόχρονα η νομοθεσία Α (που προφανώς έχει δημιουργηθεί και δημοσιευτεί πριν από τη Β) να αλλάζει τη νομοθεσία Β.

Το ELIgr έχει IRI <http://legislation.ntua.gr/eli>. Η μορφή που παίρνει το βασικό IRI της κάθε πηγής νομοθεσίας βασίζεται στο πρότυπο ELI και είναι η εξής:

`http://legislation.ntua.gr/eli/{type}/{year}/{number}/`,

όπου type ο τύπος της πηγής νομοθεσίας (Νόμος, Προεδρικό Διάταγμα κ.λ.π.), year το έτος δημιουργίας της και number ο αύξων αριθμός της. Για την αναφορά σε συγκεκριμένο τμήμα της πηγής νομοθεσίας, το IRI, πάλι σύμφωνα με το πρότυπο ELI, παίρνει τη μορφή:

`http://legislation.ntua.gr/eli/{type}/{year}/{number}/chapter/{chapter number}/article/{article number}/paragraph/{paragraph number}/`,

όπου τα chapter (κεφάλαιο), article (άρθρο) και paragraph (παράγραφος) και οι αριθμοί τους μπορούν και να παραλειφθούν ανάλογα με την περίπτωση (π.χ. αν η πηγή νομοθεσίας είναι χωρισμένη μόνο σε άρθρα και όχι και σε κεφάλαια, ή αν θέλουμε να αναφερθούμε σε ένα ολόκληρο άρθρο και όχι στις επιμέρους παραγράφους του). Σε περίπτωση που θέλουμε να αναφερθούμε στην εισαγωγή (Introduction), το Εξώφυλλο (Cover) ή τον Επίλογο (Conclusion) της νομοθεσίας, χρησιμοποιούμε, μετά το {number}/, τις συντομογραφίες intro, cover και conc αντίστοιχα.

Για τον κάθε δυνατό τύπο (type) του νομικού εγγράφου, έχει οριστεί μια συντομογραφία, η οποία χρησιμοποιείται στο IRI. Η λίστα με τις συντομογραφίες αυτές είναι η εξής:

Τύπος Νομικού Εγγράφου	Συντομογραφία
Τυπικός Νόμος (Act)	act
Διοικητική Πράξη (Administrative_Act)	aact
Τροπολογία (Amendment)	amd
Σύνταγμα (Constitution)	const
Απόφαση (Decision)	dec
Οδηγία/Εγκύκλιος (Instruction)	inst
Διεθνής Σύμβαση (International_Convention)	intconv
Νομοθετική Πράξη (Legislative_Act)	lact
Προεδρικό Διάταγμα (Presidential_Decree)	pd
Κανονισμός (Regulation)	reg

Πίνακας 2: Συντομογραφίες Τύπων Νομικών Εγγράφων στο ELIgr

Σημειώνουμε πως σε όλες τις περιπτώσεις έχουν επιλεγθεί οι αγγλικοί όροι ως βάση για τις συντομογραφίες, έτσι ώστε να διευκολύνεται η επικοινωνία μεταξύ διεθνών συστημάτων που χρησιμοποιούν Διασυνδεδεμένα Δεδομένα, αλλά και η κατανόηση των όρων από χρήστες που δε γνωρίζουν ελληνικά.

3.2 Παράδειγμα Εφαρμογής

Στην περούσα Ενότητα θα χρησιμοποιήσουμε το Νόμο 3918/2011 ως παράδειγμα για να παρουσιάσουμε το πώς μοντελοποιείται ένα νομικό έγγραφο σε μορφή RDF σημασιολογικού εγγράφου με βάση την οντολογία ELIgr που περιγράψαμε παραπάνω. Σημειώνεται πως θα γίνουν ορισμένες αναφορές σε χαρακτηριστικά των LegalDocML εγγράφων από την εργασία [52], τη μορφή των οποίων θα αναλύσουμε εκτενέστερα στην Ενότητα 3.3.

Θα ξεκινήσουμε δημιουργώντας το βασικό IRI του συγκεκριμένου νόμου το οποίο, με βάση όσα αναφέραμε νωρίτερα, παίρνει την εξής μορφή:

<http://legislation.ntua.gr/eli/act/2011/3918/>

Θα συνεχίσουμε με τη δομή των RDF σημασιολογικών εγγράφων που δημιουργήσαμε, η οποία αποτελείται από τα εξής τμήματα:

- **Βασικές πληροφορίες της Πηγής Νομοθεσίας:** Εδώ δίνονται πληροφορίες που αφορούν τα μεταδεδομένα της πηγής νομοθεσίας. Τέτοιες πληροφορίες είναι το IRI της, οι ημερομηνίες έκδοσης και εφαρμογής της, το μοναδικό αναγνωριστικό της (`id_local`), τα άτομα που την υπέγραψαν, τον εκδότη της (δηλαδή, για τα ελληνικά δεδομένα, το Εθνικό Τυπογραφείο), τον τίτλο της και τον τύπο της.

```
<eli:LegalResource rdf:about="http://legislation.ntua.gr/eli/act/2011/3918/">
  <eli:first_date_entry_in_force rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2011-03-02</eli:first_date_entry_in_force>
  <eli:date_publication rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2011-03-02</eli:date_publication>
  <eli:id_local>3918/ΠΡΩΤΟ/2011-03-02</eli:id_local>
  <eli:passed_by rdf:resource="http://legislation.ntua.gr/eli#Πρόεδρος_Δημοκρατίας"/>
  <eli:passed_by rdf:resource="http://legislation.ntua.gr/eli#Υπουργός"/>
  <eli:is_published_by rdf:resource="http://legislation.ntua.gr/eli#Εθνικό_Τυπογραφείο"/>
  <eli:title xml:lang="gr">Διαρθρωτικές αλλαγές στο σύστημα υγείας και άλλες διατάξεις.</eli:title>
  <eli:ResourceType rdf:resource="http://legislation.ntua.gr/eli#Legislative Act"/>
```

Εικόνα 13: Βασικές Πληροφορίες της Πηγής Νομοθεσίας στο RDF αρχείο

Σημειώνεται πως το `id_local` αποτελείται από τον αριθμό του νόμου, τον αριθμό του τεύχους σε όποια μορφή αυτός υπάρχει στο LegalDocML αρχείο (είτε ολογράφως είτε με αριθμό) και την ημερομηνία έκδοσης του νόμου, χωρισμένα με `</>`. Στις περιπτώσεις αποφάσεων ή εγκυκλίων, επειδή δεν υπάρχει ΦΕΚ, το `id_local` έχει τη μορφή που έχει τη τιμή `number` της κλάσης `publication` στα LegalDocML αρχεία.

Μια πληροφορία που δεν προκύπτει ως μεταδεδομένο από τα LegalDocML αρχεία είναι οι υπογράφωντες τη νομοθεσία. Αυτό θα μπορούσε να επιτευχθεί μόνο με κατάλληλο parsing στον επίλογο της κάθε νομοθεσίας. Έτσι, όταν αναφερόμαστε σε αυτά τα άτομα στο RDF αρχείο, αναφέρουμε μόνο τις κλάσεις στις οποίες ανήκουν (Υπουργός, Πρόεδρος_Δημοκρατίας κ.λ.π.).

- **Περιεχόμενα της Πηγής Νομοθεσίας:** Εδώ βρίσκεται η λίστα με όλα τα άρθρα (ή κεφάλαια, αν υπάρχουν), στα οποία χωρίζεται η πηγή νομοθεσίας. Η λίστα αυτή είναι άλλο ένα «τμήμα» που αφορά τα μεταδεδομένα της πηγής νομοθεσίας. Πάντα περιλαμβάνονται σε αυτή οι ενότητες εξώφυλλο (`cover`), εισαγωγή (`intro`) και επίλογος (`concl`). Στα εκάστοτε τμήματα, στη συνέχεια, σημειώνεται το είδος τους (εξώφυλλο, εισαγωγή, επίλογος, κεφάλαιο, άρθρο ή παράγραφος) και ο τίτλος τους, αν αυτός υπάρχει. Στην περίπτωση που κάποιο άρθρο είναι χωρισμένο σε παραγράφους (ή κάποιο κεφάλαιο σε άρθρα και μετά σε παραγράφους), σημειώνονται και αυτά κάτω από το άρθρο (ή κεφάλαιο) με τον ίδιο τρόπο.

```

<eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/cover"/>
<eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/intro"/>
<eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1"/>
<eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/2"/>
<eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/3"/>
<eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/4"/>
<eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/5"/>
<eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/6"/>
<eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/7"/>
<eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/8"/>
<eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/conc"/>

```

Εικόνα 14: Περιεχόμενα Πηγής Νομοθεσίας (Λίστα Κεφαλαίων) στο RDF αρχείο

```

<eli:LegalResource rdf:about="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/">
  <eli:LegalResourceSubdivision rdf:resource="http://legislation.ntua.gr/eli#Chapter"/>
  <eli:number>ΚΕΦΑΛΑΙΟ Α'</eli:number>
  <eli:subdivision_title xmlns="http://legislation.ntua.gr/eli#">ΕΥΕΘΗΜΑ ΠΡΟΜΗΘΕΙΩΝ ΥΓΕΙΑΣ</eli:subdivision_title>
  <eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/article/1/">
  <eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/article/2/">
  <eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/article/3/">
  <eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/article/4/">
  <eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/article/5/">
  <eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/article/6/">
  <eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/article/7/">
  <eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/article/8/">
  <eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/article/9/">
  <eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/article/10/">
  <eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/article/11/">
  <eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/article/12/">
  <eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/article/13/">
  <eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/article/14/">
  <eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/article/15/">
  <eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/article/16/">
</eli:LegalResource>

```

Εικόνα 15: Περιεχόμενα Πηγής Νομοθεσίας (Λίστα Άρθρων Κεφαλαίου) στο RDF αρχείο

```

<eli:LegalResource rdf:about="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/article/1/">
  <eli:LegalResourceSubdivision rdf:resource="http://legislation.ntua.gr/eli#Article"/>
  <eli:number>Άρθρο 1</eli:number>
  <eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/article/1/paragraph/1/">
  <eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/article/1/paragraph/2/">
  <eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/article/1/paragraph/3/">
  <eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/article/1/paragraph/4/">
  <eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/article/1/paragraph/5/">
  <eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/article/1/paragraph/6/">
  <eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/article/1/paragraph/7/">
  <eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/article/1/paragraph/8/">
  <eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/article/1/paragraph/9/">
  <eli:has_part rdf:resource="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/article/1/paragraph/10/">
</eli:LegalResource>

```

Εικόνα 16: Περιεχόμενα Πηγής Νομοθεσίας (Λίστα Παραγράφων Άρθρου) στο RDF αρχείο

Σημειώνεται πως σε κάθε κεφάλαιο, άρθρο ή παράγραφο νόμου προστέθηκε το eli:number, το οποίο είναι ένα string που αναφέρει τον αριθμό του εκάστοτε τμήματος όπως αναφέρεται στο νόμο (π.χ. "ΚΕΦΑΛΑΙΟ ΔΕΥΤΕΡΟ", "Άρθρο 15", "ΠΑΡΑΓΡΑΦΟΣ 1" κ.λ.π.). Αυτό έγινε διότι καμιά φορά η αρίθμηση (των άρθρων, κυρίως) είτε έχει γίνει λάθος, είτε δε γίνεται με νούμερο αλλά ολογράφως, είτε είναι διαφορετική από αυτήν που έχουμε κατά τη δημιουργία των IRI. Συγκεκριμένα, στα IRI, όταν έχουμε κεφάλαια, η αρίθμηση έχει τη μορφή "chapter/1/article1, chapter/1/article/2, chapter/2/article/1...", ενώ στο νόμο το chapter/2/article/1 αναφέρεται συνήθως ως Άρθρο 3, όντας το 3^ο άρθρο στη σειρά.

- Αναφορές (Citations) σε άλλες Πηγές Νομοθεσίας: Αποτελεί το τελευταίο «τμήμα» μεταδεδομένων. Είναι μια λίστα με τις υπόλοιπες πηγές νομοθεσίας που αναφέρονται στην πηγή νομοθεσίας που περιγράφεται. Σημειώνεται πως στα LegalDocML αρχεία με τα οποία δουλεύουμε, κάποιες από τις αναφορές σε άλλες πηγές νομοθεσίας δεν αναφέρονται ως δεδομένο, αλλά μπορούν να προκύψουν μόνο με κατάλληλη ανάλυση κειμένου. Δεδομένου πως η ανάλυση αυτή είναι πέρα από τους σκοπούς της παρούσας εργασίας, στα RDF αρχεία «φαίνονται» ως αναφορές μόνο όσες αναφορές «φαίνονται» και στα LegalDocML αρχεία. Επιπλέον, σε κάποια RDF αρχεία είναι δυνατόν μια αναφορά στην ίδια πηγή νομοθεσίας να υπάρχει παραπάνω από μια φορές. Αυτό συμβαίνει διότι η συγκεκριμένη πηγή νομοθεσίας αναφέρεται παραπάνω από μια φορές και στην πηγή νομοθεσίας που περιγράφεται.

```

<eli:cites rdf:resource="http://legislation.ntua.gr/eli/pd/2007/60"/>
<eli:cites rdf:resource="http://legislation.ntua.gr/eli/pd/2000/95"/>
<eli:cites rdf:resource="http://legislation.ntua.gr/eli/act/1999/2725"/>
<eli:cites rdf:resource="http://legislation.ntua.gr/eli/act/2005/3370"/>
<eli:cites rdf:resource="http://legislation.ntua.gr/eli/pd/2000/95"/>
<eli:cites rdf:resource="http://legislation.ntua.gr/eli/pd/2000/95"/>
<eli:cites rdf:resource="http://legislation.ntua.gr/eli/pd/2000/95"/>
<eli:cites rdf:resource="http://legislation.ntua.gr/eli/act/2004/3252"/>
<eli:cites rdf:resource="http://legislation.ntua.gr/eli/act/1997/2519"/>
<eli:cites rdf:resource="http://legislation.ntua.gr/eli/act/2010/3852"/>
<eli:cites rdf:resource="http://legislation.ntua.gr/eli/act/2005/3402"/>
<eli:cites rdf:resource="http://legislation.ntua.gr/eli/pd/2001/52"/>
<eli:cites rdf:resource="http://legislation.ntua.gr/eli/pd/2001/52"/>
<eli:cites rdf:resource="http://legislation.ntua.gr/eli/pd/2001/52"/>
<eli:cites rdf:resource="http://legislation.ntua.gr/eli/pd/2001/52"/>
<eli:cites rdf:resource="http://legislation.ntua.gr/eli/pd/2001/52"/>
<eli:cites rdf:resource="http://legislation.ntua.gr/eli/pd/2001/52"/>
<eli:cites rdf:resource="http://legislation.ntua.gr/eli/pd/2001/52"/>
<eli:cites rdf:resource="http://legislation.ntua.gr/eli/pd/2001/52"/>
<eli:cites rdf:resource="http://legislation.ntua.gr/eli/pd/2001/52"/>
<eli:cites rdf:resource="http://legislation.ntua.gr/eli/pd/2001/52"/>
<eli:cites rdf:resource="http://legislation.ntua.gr/eli/act/2008/3730"/>

```

Εικόνα 17: Αναφορές (Citations) σε άλλες Πηγές Νομοθεσίας στο RDF αρχείο

- Περιεχόμενο της πηγής Νομοθεσίας: Εδώ παρουσιάζεται το κείμενο της κάθε ενότητας (εξώφυλλο, εισαγωγή, επίλογος, κεφάλαιο, άρθρο ή παράγραφος) της πηγής νομοθεσίας, εφόσον αυτό υπάρχει (και δεν είναι χωρισμένο σε άλλα τμήματα όπως, για παράδειγμα, στην περίπτωση ενός άρθρου χωρισμένου παραγράφους). Αναφέρεται, επίσης, η γλώσσα στην οποία είναι γραμμένο το κείμενο, ο τύπος της ενότητας, καθώς και ο τίτλος της, αν αυτός υπάρχει. Στα παραδείγματα που ακολουθούν, τμήματα του κειμένου έχουν περικυβεί για λόγους ευκρίνειας.

```

<eli:LegalResource rdf:about="http://legislation.ntua.gr/eli/act/2011/3918/cover/">
  <eli:LegalResourceSubdivision rdf:resource="http://legislation.ntua.gr/eli#Cover"/>
  <text xmlns="http://legislation.ntua.gr/eli#" xml:lang="gr">ΕΦΗΜΕΡΙΣ ΤΗΣ ΚΥΒΕΡΝΗΣΕΩΣ
ΤΗΣ ΕΛΛΗΝΙΚΗΣ ΔΗΜΟΚΡΑΤΙΑΣ
ΤΕΥΧΟΣ ΠΡΩΤΟ Αρ. Φύλλου 31
2 Μαρτίου 2011
</text>
</eli:LegalResource>

```

Εικόνα 18: Περιεχόμενο Εξωφύλλου Πηγής Νομοθεσίας στο RDF αρχείο

```

<eli:LegalResource rdf:about="http://legislation.ntua.gr/eli/act/2011/3918/intro/">
  <eli:LegalResourceSubdivision rdf:resource="
  http://legislation.ntua.gr/eli#Introduction"/>
  <text xmlns="http://legislation.ntua.gr/eli#" xml:lang="gr">
  Εκδίδουμε τον ακόλουθο νόμο που ψήφισε η Βουλή:</text>
</eli:LegalResource>

```

Εικόνα 19: Περιεχόμενο Εισαγωγής Πηγής Νομοθεσίας στο RDF αρχείο

```

<eli:LegalResource rdf:about="http://legislation.ntua.gr/eli/act/2011/3918/chapter/1/article/14/">
  <eli:LegalResourceSubdivision rdf:resource="http://legislation.ntua.gr/eli#Article"/>
  <text xmlns="http://legislation.ntua.gr/eli#" xml:lang="gr">Οι τιμές για την προμήθεια
  ιατροτεχνολογικών προϊόντων, που κατ' εφαρμογή της διάταξης της παραγράφου 12 του άρθρου 27
  του ν. 3867/2010 (ΦΕΚ 128 Α΄) εναρμονίστηκαν με τις χαμηλότερες τιμές της εγχώριας αγοράς,
  όπως καταγράφονται στο Παρατηρητήριο Τιμών, που θεσπίστηκε με το άρθρο 24 του ν. 3846/2010
  (ΦΕΚ 66 Α΄), αναπροσαρμόζονται, από 1.1.2010, σύμφωνα με τις τιμές του ως άνω Παρατηρητηρίου
  Τιμών. Η μείωση του ποσού των ενταλαμάτων προμήθειας που αφορούν τα ανωτέρω προϊόντα θα γίνεται
  με ισόποση μείωση έκπτωσης επί των ενταλαμάτων, που θα εκδοθούν και θα αφορούν πληρωμές προμηθευτών
  των ως άνω προϊόντων, του τελευταίου πενταμήνου του έτους 2010.</text>
</eli:LegalResource>

```

Εικόνα 20: Περιεχόμενο Άρθρου Πηγής Νομοθεσίας στο RDF αρχείο

```

<eli:LegalResource rdf:about="http://legislation.ntua.gr/eli/act/2011/3918/conc/">
  <eli:LegalResourceSubdivision rdf:resource="http://legislation.ntua.gr/eli#Conclusion"/>
  <text xmlns="http://legislation.ntua.gr/eli#" xml:lang="gr">Αθήνα, 22 Φεβρουαρίου 2011
  Ο ΠΡΟΕΔΡΟΣ ΤΗΣ ΔΗΜΟΚΡΑΤΙΑΣ
  ΚΑΡΟΛΟΣ ΠΑΠΟΥΛΙΑΣ
  ΟΙ ΥΠΟΥΡΓΟΙ
  ΕΣΩΤΕΡΙΚΩΝ, ΑΠΟΚΕΝΤΡΩΣΗΣ ΟΙΚΟΝΟΜΙΚΩΝ ΟΙΚΟΝΟΜΙΑΣ,
  ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΗΣ ΔΙΑΚΥΒΕΡΝΗΣΗΣ ΑΝΤΑΓΩΝΙΣΤΙΚΟΤΗΤΑΣ ΚΑΙ ΝΑΥΤΙΛΙΑΣ
  ΙΩΑΝΝΗΣ ΡΑΓΚΟΥΣΗΣ ΓΕΩΡΓΙΟΣ ΠΑΠΑΚΩΝΣΤΑΝΤΙΝΟΥ ΜΙΧΑΗΛ ΧΡΥΣΟΧΟΪΔΗΣ
  ΠΑΙΔΕΙΑΣ, ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗΣ ΥΓΕΙΑΣ ΑΝΑΠΛΗΡΩΤΗΣ ΕΡΓΑΣΙΑΣ
  ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ ΚΑΙ ΚΟΙΝΩΝΙΚΗΣ ΑΛΛΗΛΕΓΓΥΗΣ ΚΑΙ ΚΟΙΝΩΝΙΚΗΣ ΑΣΦΑΛΙΣΗΣ
  ΑΝΝΑ ΔΙΑΜΑΝΤΟΠΟΥΛΟΥ ΑΝΔΡΕΑΣ ΛΟΒΕΡΔΟΣ ΓΕΩΡΓΙΟΣ ΚΟΥΤΡΟΥΜΑΝΗΣ
  Θεωρήθηκε και τέθηκε η Μεγάλη Σφραγίδα του Κράτους
  Αθήνα, 22 Φεβρουαρίου 2011
  Ο ΕΠΙ ΤΗΣ ΔΙΚΑΙΟΣΥΝΗΣ ΥΠΟΥΡΓΟΣ
  Χ. ΚΑΣΤΑΝΙΔΗΣ</text>
</eli:LegalResource>

```

Εικόνα 21: Περιεχόμενο Επιλόγου Πηγής Νομοθεσίας στο RDF αρχείο

4

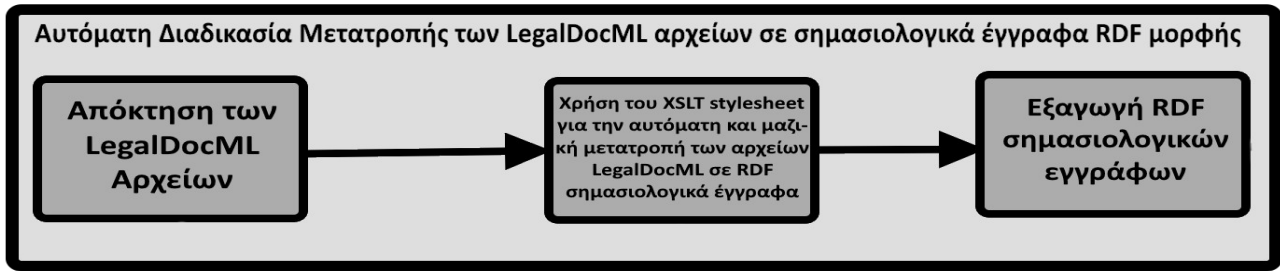
Σημασιολογικά Έγγραφα στο ELIgr

4.1 Διαδικασία Αυτόματης Μετατροπής των Σημασιολογικών Εγγράφων από το Πρότυπο LegalDocML στο Πρότυπο RDF

Για την υλοποίηση της παρούσας εργασίας, ήταν διαθέσιμο μέρος της ελληνικής νομοθεσίας σε μορφή LegalDocML, η οποία αποτελεί δομημένη μορφή XML. Αφού μελετήσαμε τα ιδιαίτερα χαρακτηριστικά της δομής των LegalDocML σημασιολογικών εγγράφων, καθώς και τα μεταδεδομένα τα οποία αυτά παρέχουν, μετατρέψαμε χειροκίνητα μερικά από αυτά σε σημασιολογικά έγγραφα που ακολουθούν το πρότυπο ELIgr. Με τον τρόπο αυτό, είδαμε στην πράξη τις απαιτήσεις και τους χειρισμούς που απαιτεί μια τέτοια μετατροπή, και καταλήξαμε στην τελική μορφή που θέλαμε να έχουν τα παραγόμενα RDF σημασιολογικά έγγραφα και που παρουσιάστηκε εκτενώς στην Ενότητα 3.2.

Στη συνέχεια, για την αυτοματοποίηση της εν λόγω μετατροπής, δημιουργήσαμε και χρησιμοποιήσαμε ένα κατάλληλο XSLT stylesheet. Το stylesheet αυτό παίρνει ως είσοδο ένα αρχείο XML σε πρότυπο LegalDocML και, λαμβάνοντας τις κατάλληλες πληροφορίες από αυτό, παράγει το αντίστοιχο XML αρχείο σε σύνταξη RDF/XML ως έξοδο.

Η συνολική διαδικασία που ακολουθείται, πλέον, για τη μετατροπή των LegalDocML αρχείων σε RDF σημασιολογικά έγγραφα με βάση την οντολογία ELIgr μπορεί να απεικονιστεί στο παρακάτω διάγραμμα ροής:



Εικόνα 22: Διάγραμμα Ροής Διαδικασίας Μετατροπής των LegalDocML σημασιολογικών εγγράφων σε RDF

Στο πρώτο βήμα του σχήματος, συγκεντρώνουμε το σύνολο των LegalDocML αρχείων που θέλουμε να μετατρέψουμε, η δομή των οποίων αναλύεται στην Ενότητα 4.2. Στο δεύτερο βήμα μετατρέπουμε όλα τα αποθηκευμένα LegalDocML αρχεία σε RDF σημασιολογικά έγγραφα με τη χρήση του XSLT Stylesheet, η λειτουργία του οποίου αναλύεται στην Ενότητα 4.3, σε σημασιολογικά έγγραφα που ακολουθούν το πρότυπο RDF (βήμα 3). Η δομή των RDF σημασιολογικών εγγράφων έχει παρουσιαστεί στην Ενότητα 3.2.

4.2 Περιγραφή της Δομής των LegalDocML Σημασιολογικών Εγγράφων

Τα LegalDocML αρχεία που χρησιμοποιούνται στην παρούσα εργασία είναι εκείνα που παράχθηκαν από το σύστημα που αναπτύχθηκε στη διπλωματική εργασία [52] της Σχολής Θετικών Επιστημών και Τεχνολογίας του Ελληνικού Ανοικτού Πανεπιστημίου. Πρόκειται για 884 XML αρχεία τα οποία ακολουθούν το πρότυπο LegalDocML (πρόην AkomaNtoso) και περιγράφουν κυρίως ελληνικούς Νόμους και Προεδρικά Διατάγματα, ενώ λίγα από αυτά περιγράφουν Εγκυκλίους και Αποφάσεις. Σημειώνεται πως κάποιοι Νόμοι που μας δόθηκαν είναι Τροποποιήσεις (Amendments) άλλων νόμων, όμως επειδή στα μεταδεδομένα του LegalDocML αρχείου δηλώνονται ως Νόμοι, τους έχουμε εντάξει στην κατηγορία των Νόμων και τους επεξεργαζόμαστε όπως τους Νόμους, καθώς η δήλωσή τους ως Τροποποιήσεις θα απαιτούσε κατάλληλη ανάλυση κειμένου (parsing), κάτι που ξεπερνά τους σκοπούς της παρούσας εργασίας.

Η δομή των LegalDocML αρχείων αυτών έχει, σε γενικές γραμμές, τις εξής ενότητες (ως παράδειγμα χρησιμοποιείται ξανά ο Νόμος 3918/2011):

- **meta:** Σε αυτό το κομμάτι περιγράφονται τα μεταδεδομένα του αρχείου, δηλαδή στοιχεία όπως η γλώσσα της πηγής νομοθεσίας, η ημέρα δημοσίευσής της, ο τίτλος της, το είδος της, κ.λ.π..

```

<meta>
  <identification source="#HB_bot">
    <FRBRWork>
      <FRBRthis value="/gr/act/2011/3918/">
      <FRBRuri value="/gr/act/2011/3918/">
      <FRBRdate date="2011-03-02" name="Publish"/>
      <FRBRauthor as="#author" href="#"/>
      <FRBRcountry value="gr"/>
    </FRBRWork>
    <FRBRExpression>
      <FRBRthis value="/gr/act/2011/3918/3918_2011/">
      <FRBRuri value="/gr/act/2011/3918/3918_2011/">
      <FRBRdate date="2015-10-05" name="INSERT"/>
      <FRBRauthor as="#editor" href="#HB_bot"/>
      <FRBRlanguage language="gr"/>
    </FRBRExpression>
    <FRBRManifestation>
      <FRBRthis value="/gr/act/2011/3918/3918_2011.xml/">
      <FRBRuri value="/gr/act/2011/3918/3918_2011.akn/">
      <FRBRdate date="2015-10-05" name="XMLConversion"/>
      <FRBRauthor as="#editor" href="#HB_bot"/>
    </FRBRManifestation>
  </identification>
  <publication date="2011-03-02" number="ΤΕΥΧΟΣ ΠΡΩΤΟ Αρ. φύλλου 31"
  name="et" showAs="ΕΘΝΙΚΟ ΤΥΠΟΓΡΑΦΕΙΟ"/>
  <classification source="#main">
    <keyword dictionary="topK" showAs="παραγραφ" value="παραγραφ"/>
    <keyword dictionary="topK" showAs="αρθρ" value="αρθρ"/>
    <keyword dictionary="topK" showAs="υγει" value="υγει"/>
    <keyword dictionary="topK" showAs="κοινωνικ" value="κοινωνικ"/>
    <keyword dictionary="topK" showAs="υπαλληλ" value="υπαλληλ"/>
    <keyword dictionary="topK" showAs="αποφασ" value="αποφασ"/>
    <keyword dictionary="topK" showAs="υπηρεσ" value="υπηρεσ"/>
    <keyword dictionary="topK" showAs="αλληλεγγυ" value="αλληλεγγυ"/>
    <keyword dictionary="topK" showAs="ιατρ" value="ιατρ"/>
    <keyword dictionary="topK" showAs="υπουργ" value="υπουργ"/>
  </classification>
  <classification source="#διδυμια"/>

```

Εικόνα 24: Μεταδεδομένα LegalDocML Αρχείου (μέρος 1ο)

```

  <eventRef type="generation" date="2011-03-02"/>
</lifecycle>
<references source="#HB_bot">
  <original href="/gr/act/2011/3918/files/3918_2011.txt" showAs="Πρωτότυπο"/>
</references>
<proprietary source="#HB_bot">
  <DCTerms xmlns:ns2="http://docs.oasis-open.org/legaldocml/ns/akn/3.0/CSD13"
  xmlns="" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
    <docIssuerRank>ΠΡΟΕΔΡΟΣ ΤΗΣ ΕΛΛΗΝΙΚΗΣ ΔΗΜΟΚΡΑΤΙΑΣ|ΥΠΟΥΡΓΟΙ</docIssuerRank>
    <dc:description>ΝΟΜΟΣ_3918_2011</dc:description>
    <dc:language>gr</dc:language>
    <dc:creator>Ο ΠΡΟΕΔΡΟΣ ΤΗΣ ΕΛΛΗΝΙΚΗΣ ΔΗΜΟΚΡΑΤΙΑΣ</dc:creator>
    <dcterms:created>2011-03-02</dcterms:created>
    <dc:format>text/xml</dc:format>
    <dc:type2>ACT</dc:type2>
    <dc:type>ΝΟΜΟΣ</dc:type>
    <docKYA/>
    <dc:title>Διαρθρωτικές αλλαγές στο σύστημα υγείας και άλλες διατάξεις.</dc:title>
    <docProtocol/>
    <docPol/>
    <dc:identifier>/gr/act/2011/3918/3918_2011.akn</dc:identifier>
    <dc:source>/gr/act/2011/3918/files/3918_2011.txt</dc:source>
    <dc:publisher>ΕΘΝΙΚΟ ΤΥΠΟΓΡΑΦΕΙΟ</dc:publisher>
  </DCTerms>
</proprietary>
</meta>

```

Εικόνα 23: Μεταδεδομένα LegalDocML Αρχείου (μέρος 2ο)

- coverPage και preface: Σε αυτές τις ενότητες παρουσιάζεται το εξώφυλλο της πηγής νομοθεσίας, καθώς επίσης και η δομή της, δηλαδή ο αριθμός και η σειρά των κεφαλαίων, των άρθρων και των παραγράφων που την αποτελούν. Επειδή ο Νόμος 3918/2011 περιέχει πάρα πολλά κεφάλαια, άρθρα και παραγράφους, κάποια από αυτά έχουν παραληφθεί στις ακόλουθες εικόνες για λόγους ευκρίνειας.

```

<coverPage>
  <p>ΕΦΗΜΕΡΙΣ ΤΗΣ ΚΥΒΕΡΝΗΣΕΩΣ
ΤΗΣ ΕΛΛΗΝΙΚΗΣ ΔΗΜΟΚΡΑΤΙΑΣ
ΤΕΥΧΟΣ ΠΡΩΤΟ Αρ. φύλλου 31
2 Μαρτίου 2011
</p>

```

Εικόνα 25: Εξώφυλλο και Δομή της Πηγής Νομοθεσίας στο LegalDocML Αρχείο (μέρος 1ο)

```

<toc eId="toc" wId="/gr/act/2011/3918/main/toc/">
  <tocItem level="1" href="/gr/act/2011/3918/main/chp/1/">ΚΕΦΑΛΑΙΟ Α΄ - ΣΥΣΤΗΜΑ ΠΡΟΜΗΘΕΙΩΝ ΥΓΕΙΑΣ</tocItem>
  <tocItem level="2" href="/gr/act/2011/3918/main/chp/1/art/1/">Άρθρο 1 - Πεδίο εφαρμογής - Ορισμοί</tocItem>
  <tocItem level="3" href="/gr/act/2011/3918/main/chp/1/art/1/para/1/">ΠΑΡΑΓΡΑΦΟΣ 1 </tocItem>
  <tocItem level="3" href="/gr/act/2011/3918/main/chp/1/art/1/para/2/">ΠΑΡΑΓΡΑΦΟΣ 2 </tocItem>
  <tocItem level="3" href="/gr/act/2011/3918/main/chp/1/art/1/para/3/">ΠΑΡΑΓΡΑΦΟΣ 3 </tocItem>
  <tocItem level="3" href="/gr/act/2011/3918/main/chp/1/art/1/para/4/">ΠΑΡΑΓΡΑΦΟΣ 4 </tocItem>
  <tocItem level="3" href="/gr/act/2011/3918/main/chp/1/art/1/para/5/">ΠΑΡΑΓΡΑΦΟΣ 5 </tocItem>
  <tocItem level="3" href="/gr/act/2011/3918/main/chp/1/art/1/para/6/">ΠΑΡΑΓΡΑΦΟΣ 6 </tocItem>
  <tocItem level="3" href="/gr/act/2011/3918/main/chp/1/art/1/para/7/">ΠΑΡΑΓΡΑΦΟΣ 7 </tocItem>
  <tocItem level="3" href="/gr/act/2011/3918/main/chp/1/art/1/para/8/">ΠΑΡΑΓΡΑΦΟΣ 8 </tocItem>
  <tocItem level="3" href="/gr/act/2011/3918/main/chp/1/art/1/para/9/">ΠΑΡΑΓΡΑΦΟΣ 9 </tocItem>
  <tocItem level="3" href="/gr/act/2011/3918/main/chp/1/art/1/para/10/">ΠΑΡΑΓΡΑΦΟΣ 10 </tocItem>
  <tocItem level="2" href="/gr/act/2011/3918/main/chp/1/art/2/">Άρθρο 2 - Αντικείμενο συμβάσεων προμηθειών</tocItem>
  <tocItem level="3" href="/gr/act/2011/3918/main/chp/1/art/2/para/1/">ΠΑΡΑΓΡΑΦΟΣ 1 </tocItem>
  <tocItem level="3" href="/gr/act/2011/3918/main/chp/1/art/2/para/2/">ΠΑΡΑΓΡΑΦΟΣ 2 </tocItem>

```

Εικόνα 26: Εξώφυλλο και Δομή της Πηγής Νομοθεσίας στο LegalDocML Αρχείο (μέρος 2ο)

```

</toc>
</coverPage>
<preface>
  <p>ΝΟΜΟΣ ΥΠ' ΑΡΙΘ. 3918
Διαρθρωτικές αλλαγές στο σύστημα υγείας
και άλλες διατάξεις.
Ο ΠΡΟΕΔΡΟΣ
ΤΗΣ ΕΛΛΗΝΙΚΗΣ ΔΗΜΟΚΡΑΤΙΑΣ</p>
</preface>

```

Εικόνα 27: Εξώφυλλο και Δομή της Πηγής Νομοθεσίας στο LegalDocML Αρχείο (μέρος 3ο)

- preamble: Το εισαγωγικό κείμενο της πηγής νομοθεσίας.

```

<preamble>
  <p>Εκδίδομε τον ακόλουθο νόμο που ψήφισε η Βουλή:</p>
</preamble>

```

Εικόνα 28: Εισαγωγικό Κείμενο της Πηγής Νομοθεσίας στο LegalDocML Αρχείο

- **body:** Ο τίτλος και το κείμενο κάθε κεφαλαίου, άρθρου και παραγράφου από τα οποία αποτελείται η πηγή της νομοθεσίας. Είναι, ουσιαστικά, το περιεχόμενο της πηγής νομοθεσίας. Επειδή ο Νόμος 3918/2011 περιέχει πάρα πολλά κεφάλαια, άρθρα και παραγράφους, κάποια από αυτά έχουν παραληφθεί στην εικόνα για λόγους ευκρίνειας.

```

<body eId="/gr/act/2011/3918/main/" wId="/gr/act/2011/3918/main/">
  <chapter eId="chp/1/" wId="/gr/act/2011/3918/main/chp/1/">
    <num>ΚΕΦΑΛΑΙΟ Α' </num>
    <heading eId="chp/1/heading/" wId="/gr/act/2011/3918/main/chp/1/heading/">ΣΥΣΤΗΜΑ ΠΡΟΜΗΘΕΙΩΝ ΥΓΕΙΑΣ</heading>
    <article eId="chp/1/art/1/" wId="/gr/act/2011/3918/main/chp/1/art/1/">
      <num>Άρθρο 1</num>
      <heading eId="chp/1/art/1/heading/" wId="/gr/act/2011/3918/main/chp/1/art/1/heading/">Πεδίο εφαρμογής - Ορισμοί</heading>
      <paragraph eId="chp/1/art/1/para/1/" wId="/gr/act/2011/3918/main/chp/1/art/1/para/1/">
        <num>ΠΑΡΑΓΡΑΦΟΣ 1 </num>
        <content eId="chp/1/art/1/para/1/content/" wId="/gr/act/2011/3918/main/chp/1/art/1/para/1/content/">
          <p>Στις διατάξεις του παρόντος κεφαλαίου, με την επιφύλαξη των διατάξεων του π.δ. 60/2007 (ΦΕΚ 64 Α') «Προσαρμογή της ελληνικής νομοθεσίας στις διατάξεις της Οδηγίας 2004/18/ΕΚ, όπως τροποποιήθηκε με την Οδηγία 2005/51/ΕΚ της Επιτροπής και την Οδηγία 2005/ 75/ΕΚ του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου της 16ης Νοεμβρίου 2005», υπάγονται οι συμβάσεις προμηθειών και υπηρεσιών, όπως και οι συμφωνίες-πλαίσιο για την κάλυψη των αναγκών όλων των νοσοκομείων του Εθνικού Συστήματος Υγείας (Ε.Σ.Υ.) και των αποκεντρωμένων μονάδων αυτών, καθώς και των Μονάδων Κοινωνικής Φροντίδας και των Ν.Π.Δ.Δ. που ασκούν δραστηριότητες στους τομείς υγείας και κοινωνικής αλληλεγγύης και εποπτεύονται από τον Υπουργό Υγείας και Κοινωνικής Αλληλεγγύης είτε η λειτουργία τους εποπτεύεται και ελέγχεται από τις Υγειονομικές Περιφέρειες κατά τα οριζόμενα στο άρθρο 2 του ν. 3329/2005 (ΦΕΚ 81 Α') είτε υπάγονται απευθείας στον Υπουργό Υγείας και Κοινωνικής Αλληλεγγύης. Εξαιρούνται οι περιπτώσεις των παραγράφων 3 και 4 του παρόντος άρθρου. Τις συμβάσεις προμηθειών και υπηρεσιών, όπως και τις συμφωνίες-πλαίσιο συνάπτουν οι αναθέτουσες αρχές με έναν ή περισσότερους οικονομικούς φορείς.</p>
        </content>
      </paragraph>
      <paragraph eId="chp/1/art/1/para/2/" wId="/gr/act/2011/3918/main/chp/1/art/1/para/2/">
        <num>ΠΑΡΑΓΡΑΦΟΣ 2 </num>
        <content eId="chp/1/art/1/para/2/content/" wId="/gr/act/2011/3918/main/chp/1/art/1/para/2/content/">
          <p>Με κοινή απόφαση των Υπουργών Οικονομίας, Ανταγωνιστικότητας και Ναυτιλίας, Υγείας και Κοινωνικής Αλληλεγγύης και του κατά περίπτωση αρμόδιου Υπουργού, η εφαρμογή του παρόντος μπορεί να επεκτείνεται στις συμβάσεις προμηθειών και υπηρεσιών και σε συμφωνίες-πλαίσιο για την κάλυψη των αναγκών άλλων φορέων που δραστηριοποιούνται στο χώρο της υγείας και εποπτεύονται και ελέγχονται από άλλους Υπουργούς, μεταξύ των οποίων τα νοσοκομεία ή άλλες σχετικές μονάδες που εποπτεύονται από το Υπουργείο Εθνικής Άμυνας. Με την ίδια κοινή υπουργική απόφαση προβλέπεται η συμμετοχή δύο εκπροσώπων του εκάστοτε φορέα, στον οποίο επεκτείνεται η εφαρμογή των διατάξεων του Κεφαλαίου Α', στη Συντονιστική Επιτροπή Προμηθειών του άρθρου 6, καθώς και κάθε άλλο θέμα για τη συμμόρφωση στις εν λόγω διατάξεις.</p>
        </content>
      </paragraph>
      <paragraph eId="chp/1/art/1/para/3/" wId="/gr/act/2011/3918/main/chp/1/art/1/para/3/">
        <num>ΠΑΡΑΓΡΑΦΟΣ 3 </num>
        <content eId="chp/1/art/1/para/3/content/" wId="/gr/act/2011/3918/main/chp/1/art/1/para/3/content/">
          <p>Δεν υπάγονται στις διατάξεις του παρόντος οι προμήθειες πάσης φύσεως εξοπλισμού και υπηρεσιών συντήρησης του εξοπλισμού των φορέων της παραγράφου 1 του άρθρου 1 του παρόντος που ανατίθενται στη «Δημόσια Επιχείρηση Ανέγερσης Νοσοκομειακών Μονάδων» «Δ.Ε.Π.Α.ΝΟ.Μ. Α.Ε.», σύμφωνα με τα άρθρα 4 και 5 του ν. 1398/1983 (ΦΕΚ 144 Α'), όπως συμπληρώθηκαν από το άρθρο 54 του ν. 2519/1997 (ΦΕΚ 165 Α') και την παράγραφο 3 του άρθρου 22 του ν. 2737/1999 (ΦΕΚ 174 Α').</p>
        </content>
      </paragraph>
      <paragraph eId="chp/1/art/1/para/4/" wId="/gr/act/2011/3918/main/chp/1/art/1/para/4/">
        <num>ΠΑΡΑΓΡΑΦΟΣ 4 </num>

```

Εικόνα 29: Περιεχόμενο κάθε υποενότητας της Πηγής Νομοθεσίας στο LegalDocML Αρχείο

- **conclusions:** Το κείμενο του epilόγου της πηγής νομοθεσίας.

```

<conclusions>
  <p>Αθήνα, 22 Φεβρουαρίου 2011
  Ο ΠΡΟΕΔΡΟΣ ΤΗΣ ΔΗΜΟΚΡΑΤΙΑΣ
  ΚΑΡΟΛΟΣ ΠΑΠΟΥΛΙΑΣ
  ΟΙ ΥΠΟΥΡΓΟΙ
  ΕΣΩΤΕΡΙΚΩΝ, ΑΠΟΚΕΝΤΡΩΣΗΣ ΟΙΚΟΝΟΜΙΚΩΝ ΟΙΚΟΝΟΜΙΑΣ,
  ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΗΣ ΔΙΑΚΥΒΕΡΝΗΣΗΣ ΑΝΤΑΓΩΝΙΣΤΙΚΟΤΗΤΑΣ ΚΑΙ ΝΑΥΤΙΛΙΑΣ
  ΙΩΑΝΝΗΣ ΡΑΓΚΟΥΣΗΣ ΓΕΩΡΓΙΟΣ ΠΑΠΑΚΩΝΣΤΑΝΤΙΝΟΥ ΜΙΧΑΗΛ ΧΡΥΣΟΧΟΪΔΗΣ
  ΠΑΙΔΕΙΑΣ, ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗΣ ΥΓΕΙΑΣ ΑΝΑΠΑΗΡΩΤΗΣ ΕΡΓΑΣΙΑΣ
  ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ ΚΑΙ ΚΟΙΝΩΝΙΚΗΣ ΑΛΛΗΛΕΓΓΥΗΣ ΚΑΙ ΚΟΙΝΩΝΙΚΗΣ ΑΣΦΑΛΙΣΗΣ
  ΑΝΝΑ ΔΙΑΜΑΝΤΟΠΟΥΛΟΥ ΑΝΔΡΕΑΣ ΔΟΒΕΡΔΟΣ ΓΕΩΡΓΙΟΣ ΚΟΥΤΡΟΥΜΑΝΗΣ
  Θεωρήθηκε και τέθηκε η Μεγάλη Σφραγίδα του Κράτους
  Αθήνα, 22 Φεβρουαρίου 2011
  Ο ΕΠΙ ΤΗΣ ΔΙΚΑΙΟΣΥΝΗΣ ΥΠΟΥΡΓΟΣ
  Χ. ΚΑΣΤΑΝΙΔΗΣ</p>
</conclusions>

```

Εικόνα 30: Κείμενο Επιλόγου της Πηγής Νομοθεσίας στο LegalDocML Αρχείο

4.3 Περιγραφή της Δομής και της Λειτουργίας του Μετασχηματισμού XSLT

Στην παρούσα Ενότητα, θα παρουσιάσουμε τη λειτουργία του XSLT stylesheet. Το XSLT stylesheet μετατρέπει τα σημασιολογικά έγγραφα τύπου LegalDocML (που παρουσιάστηκαν αναλυτικά στην Ενότητα 4.2) σε RDF σημασιολογικά έγγραφα (που παρουσιάστηκαν αναλυτικά στην Ενότητα 3.2), παράγοντας τα αντίστοιχα XML αρχεία. Ειδικά σημεία ενδιαφέροντος του XSLT stylesheet είναι τα εξής:

```
<xsl:template match="akomaNtoso/act"> <!--law's metadata, then cover, intro, chapters/articles/paragraphs
and outro descriptions and content-->
<!--metadata-->
<eli:LegalResource rdf:about="{ $rooturi }">
  <eli:first_date_entry_in_force rdf:datatype="http://www.w3.org/2001/XMLSchema#date">
    <xsl:value-of select="meta/identification/FRBRWork/FRBRdate/@date"/> </eli:first_date_entry_in_force>
  <eli:date_publication rdf:datatype="http://www.w3.org/2001/XMLSchema#date">
    <xsl:value-of select="meta/publication/@date"/></eli:date_publication>
  <eli:id_local><xsl:value-of select ="$localid"/></eli:id_local>
  <eli:passed_by rdf:resource="http://legislation.ntua.gr/eli#Πρόεδρος Δημοκρατίας"/>
  <eli:passed_by rdf:resource="http://legislation.ntua.gr/eli#Υπουργός"/>
  <eli:publisher_agent rdf:resource="http://legislation.ntua.gr/eli#Εθνικό Τυπογραφείο"/>
  <xsl:apply-templates select="meta/proprietary" /> <!--reach DCTerms for title-->
  <xsl:if test="$typelaw='pd'">
    <eli:ResourceType rdf:resource="http://legislation.ntua.gr/eli#Presidential Decree"/>
  </xsl:if>
  <xsl:if test="$typelaw='act'">
    <eli:ResourceType rdf:resource="http://legislation.ntua.gr/eli#Act"/>
  </xsl:if>
  <xsl:apply-templates select="//a" /> <!--search for citations-->
  <eli:has_part rdf:resource="{concat($rooturi,'cover')}" />
  <eli:has_part rdf:resource="{concat($rooturi,'intro')}" />
  <xsl:apply-templates select="coverPage"/> <!--parts of the law-->
  <eli:has_part rdf:resource="{concat($rooturi,'conc')}" />
</eli:LegalResource>
```

Εικόνα 31: Εγγραφή Μεταδεδομένων στα RDF Σημασιολογικά Έγγραφα

Το τμήμα του XSLT stylesheet που απεικονίζεται στην Εικόνα 31 επεξεργάζεται αποκλειστικά Νόμους και Προεδρικά Διατάγματα. Εδώ γίνεται η εγγραφή των μεταδεδομένων της πηγής νομοθεσίας στο RDF αρχείο. Καταγράφονται οι εξής πληροφορίες: Το βασικό URI της πηγής νομοθεσίας, η ημερομηνία έκδοσης και έναρξης εφαρμογής της, το μοναδικό αναγνωριστικό της, τα πρόσωπα που την υπέγραψαν (πάντα πρόκειται για Υπουργούς και τον Πρόεδρο της Δημοκρατίας, οπότε αναφέρουμε τις αντίστοιχες κλάσεις), ο εκδότης της (που στην Ελλάδα είναι πάντα το Εθνικό Τυπογραφείο), το είδος της, ο τίτλος της, οι αναφορές σε άλλες πηγές νομοθεσίας (αν υπάρχουν) και τα τμήματα στα οποία χωρίζεται (Εξώφυλλο, Εισαγωγή, Επίλογος και Κεφάλαια, Άρθρα ή Παράγραφοι αν υπάρχουν). Για την εύρεση των δύο τελευταίων χρησιμοποιούνται άλλα τμήματα του XSLT stylesheet που θα αναλυθούν παρακάτω. Για τις περιπτώσεις των Εγκυκλίων και των Αποφάσεων χρησιμοποιείται ένα παρομοίωτο τμήμα του XSLT stylesheet με μικρές αλλαγές, καθώς η δομή των αντίστοιχων LegalDocML αρχείων διαφέρει σε ορισμένα σημεία.

```

<xsl:template match="coverPage"> <!--parts of law-->
  <xsl:for-each select="toc/tocItem">
    <xsl:if test="@level=1">
      <xsl:variable name="testuri" select="substring-after(@href, '/main/')" /> <!-- art/1/ -->
      <xsl:variable name="typeuri" select="substring-before($testuri, '/')" /> <!-- art or chp, the type of law part -->
      <xsl:if test="$typeuri='chp'">
        <xsl:variable name="numburi" select="substring-after($testuri, '/')" /> <!-- 1/ -->
        <xsl:variable name="newuri" select="concat($rooturi, 'chapter/', $numburi)"/>
        <eli:has_part rdf:resource="{ $newuri }"/>
      </xsl:if>
      <xsl:if test="$typeuri='art'">
        <xsl:variable name="numburi" select="substring-after($testuri, '/')" /> <!-- 1/ -->
        <xsl:variable name="newuri" select="concat($rooturi, 'article/', $numburi)"/>
        <eli:has_part rdf:resource="{ $newuri }"/>
      </xsl:if>
    </xsl:if>
  </xsl:for-each>
</xsl:template>

```

Εικόνα 32: Εγγραφή των Τμημάτων της Πηγής Νομοθεσίας στα RDF Σημασιολογικά Έγγραφα

Οι πηγές νομοθεσίας συνήθως είναι χωρισμένες είτε σε Άρθρα που χωρίζονται σε Παραγράφους, είτε σε Κεφάλαια που χωρίζονται σε Άρθρα που χωρίζονται σε Παραγράφους. Το παραπάνω τμήμα του XSLT stylesheet ελέγχει (μέσω της μεταβλητής typeuri) αν το πρώτο «επίπεδο» (level=1) τμηματοποίησης της πηγής νομοθεσίας είναι τα Άρθρα ή τα Κεφάλαια και, στη συνέχεια, εγγράφει στο αρχείο RDF τη λίστα με τα URIs των Κεφαλαίων ή Άρθρων αυτών. Τα URIs αυτά προκύπτουν προσθέτοντας τα κατάλληλα στοιχεία στο βασικό URI της πηγής νομοθεσίας.

```

<xsl:template match="//a"> <!--citations-->
  <xsl:variable name="law1" select="substring-after(@href, 'gr/')" /> <!--act/2014/4275/art/5/para/1-->
  <xsl:variable name="lawtype" select="substring-before($law1, '/')" /> <!--act/pd, the law type-->
  <xsl:variable name="law2" select="substring-after($law1, '/')" /> <!--2014/4275/art/5/para/1-->
  <xsl:variable name="lawdate" select="substring-before($law2, '/')" /> <!--2014, the law date-->
  <xsl:variable name="law3" select="substring-after($law2, '/')" /> <!--4275/art/5/para/1-->
  <xsl:variable name="lawnum" select="substring-before($law3, '/')" /> <!--4275, the law number-->
  <xsl:variable name="lawrefer" select="concat('http://legislation.ntua.gr/eli/', $lawtype,
    '/', $lawdate, '/', $lawnum)" /> <!--the referenced law-->
  <eli:cites rdf:resource="{ $lawrefer }"/>
</xsl:template>

```

Εικόνα 33: Εγγραφή των Αναφορών της Πηγής Νομοθεσίας σε άλλες Πηγές Νομοθεσίας στα RDF Σημασιολογικά Έγγραφα

Το τμήμα του XSLT αρχείου της Εικόνας 36 βρίσκει τις αναφορές της πηγής νομοθεσίας σε άλλες πηγές νομοθεσίας και εγγράφει το URI τους, το οποίο προκύπτει προσθέτοντας τα κατάλληλα στοιχεία στο βασικό URI της οντολογίας στο RDF αρχείο. Κρατούμε δεδομένα μόνο για την αναφερόμενη πηγή νομοθεσίας (δηλαδή τον τύπο της, την ημερομηνία της και τον αριθμό της) και όχι για το Κεφάλαιο, Άρθρο ή Παράγραφο της στο οποίο μπορεί να γίνεται αναφορά. Αυτό συμβαίνει διότι τα URI της οντολογίας την οποία αφορούν τα LegalDocML αρχεία έχουν διαφορετική δομή από τα δικά μας, με αποτέλεσμα να χρειάζονται κατάλληλη ανάλυση κειμένου (parsing) ώστε να μας δίνουν το ακριβές τμήμα της πηγής νομοθεσίας στην οποία γίνεται η αναφορά.

```

<xsl:template match="body">
  <!--case of articles only-->
  <xsl:for-each select="..//body/article"> <!--descriptions for articles that have no chapter, aka they are directly under 'body'-->
    <xsl:variable name="artinum" select="substring-after(@eId, '/')" /> <!--1/, article number-->
    <xsl:variable name="artiuri" select="concat($rooturi, 'article/', $artinum)" /> <!-- article's uri-->
    <eli:LegalResource rdf:about="{ $artiuri }">
      <eli:LegalResourceSubdivision rdf:resource="http://legislation.ntua.gr/eli#Article"/>
      <eli:number><xsl:value-of select="num" /></eli:number>
      <eli:subdivision title xmlns="http://legislation.ntua.gr/eli#"> <xsl:value-of select="heading" /> </eli:subdivision_title>
      <xsl:for-each select="paragraph"> <!--uri for each paragraph-->
        <xsl:variable name="paranum" select="substring-after(@eId, 'para/')" /> <!--1/, paragraph number-->
        <eli:has_part rdf:resource="{concat($artiuri, 'paragraph/', $paranum)}"/>
      </xsl:for-each>
    </eli:LegalResource>
  </xsl:for-each>

  <xsl:for-each select="..//body/article"> <!--text of each article-->
    <xsl:if test="..//content/p"> <!--if we have no content for the article itself, we don't need to make a new node for this-->
      <xsl:variable name="artinum1" select="substring-after(@eId, '/')" /> <!--1/, article number-->
      <xsl:variable name="artiuri1" select="concat($rooturi, 'article/', $artinum1)" /> <!-- article's uri-->
      <eli:LegalResource rdf:about="{ $artiuri1 }">
        <eli:LegalResourceSubdivision rdf:resource="http://legislation.ntua.gr/eli#Article"/>
        <eli:subdivision title xmlns="http://legislation.ntua.gr/eli#"> <xsl:value-of select="heading" /> </eli:subdivision_title>
        <text xmlns="http://legislation.ntua.gr/eli#" xml:lang="{ $language }"> <xsl:value-of select="content/p"/></text>
      </eli:LegalResource>
    </xsl:if>
  </xsl:for-each>

  <xsl:for-each select="..//body/article/paragraph"> <!--text of each paragraph-->
    <xsl:variable name="artinum2" select="substring-after(..//@eId, 'art/')" /> <!--1/, article number-->
    <xsl:variable name="artiuri2" select="concat($rooturi, 'article/', $artinum2)" /> <!-- article's uri-->
    <xsl:variable name="artiparanum" select="substring-after(@eId, 'para/')" /> <!--1/, paragraph number-->
    <xsl:variable name="artiparauri" select="concat($artiuri2, 'paragraph/', $artiparanum)" /> <!-- paragraph's uri-->
    <eli:LegalResource rdf:about="{ $artiparauri }">
      <eli:LegalResourceSubdivision rdf:resource="http://legislation.ntua.gr/eli#Paragraph"/>
      <eli:number><xsl:value-of select="num" /></eli:number>
      <text xmlns="http://legislation.ntua.gr/eli#" xml:lang="{ $language }"> <xsl:value-of select="content/p"/></text>
    </eli:LegalResource>
  </xsl:for-each>

```

Εικόνα 34: Εγγραφή του Περιεχομένου Άρθρων και Παραγράφων στα RDF Σημασιολογικά Έγγραφα

Το τελευταίο τμήμα του XSLT stylesheet αφορά την εγγραφή του περιεχομένου της πηγής νομοθεσίας στο RDF αρχείο, και συγκεκριμένα την περίπτωση που δεν έχουμε Κεφάλαια, αλλά έχουμε μόνο Άρθρα χωρισμένα σε Παραγράφους. Για κάθε Άρθρο, βρίσκει τον αριθμό του και παράγει το URI του κάνοντας τις κατάλληλες προσθήκες στο βασικό URI της πηγής νομοθεσίας. Στη συνέχεια, εγγράφει στο RDF αρχείο τον τύπο του (δηλαδή Article), τον αριθμό του (όπως είναι γραμμένος, χωρίς αυτός απαραίτητα να συμφωνεί με τον αριθμό του Άρθρου που δηλώνεται στο URI, όπως αναφέραμε στην Ενότητα 3.2), και τον τίτλο του.

Στη συνέχεια, για κάθε Άρθρο, ελέγχεται αν υπάρχει κείμενο με το περιεχόμενό του, χωρίς αυτό να είναι χωρισμένο σε Παραγράφους. Αν υπάρχει, γράφεται στο RDF αρχείο ξεχωριστό τμήμα για κάθε τέτοιο Άρθρο, το οποίο πάλι περιλαμβάνει το URI, τον τύπο (δηλαδή Article) και τον τίτλο του, και, επιπλέον, τη γλώσσα και το περιεχόμενό του. Επιλέξαμε την επανάληψη των βασικών πληροφοριών του κάθε Άρθρου έτσι ώστε τα RDF έγγραφα να είναι ευκολότερα στην κατανόηση για τους χρήστες.

Μια αντίστοιχη διαδικασία γίνεται και για τις Παραγράφους κάθε Άρθρου. Σε κάθε Άρθρο, για κάθε Παράγραφο βρίσκουμε τον αριθμό της και παράγουμε το URI της προσθέτοντας την κατάλληλη πληροφορία στο URI του Άρθρου στο οποίο ανήκει. Στη συνέχεια, για κάθε Παράγραφο, γράφεται ξεχωριστό τμήμα στο RDF αρχείο που περιλαμβάνει το URI της, τον τύπο της (δηλαδή Paragraph), τον αριθμό της (όπως είναι γραμμένος στο κείμενο, χωρίς αυτός απαραίτητα να συμφωνεί με τον αριθμό της Παραγράφου που δηλώνεται στο URI, όπως αναφέραμε στην Ενότητα 3.2), τη γλώσσα και το περιεχόμενό της.

Έτσι, παράγονται, ουσιαστικά, τρία διαφορετικά τμήματα στο RDF έγγραφο. Το πρώτο περιλαμβάνει τα

βασικά μεταδεδομένα για το κάθε Άρθρο και τα «περιεχόμενά» του (δηλαδή τη λίστα με τα URI των Παραγράφων τις οποίες περιλαμβάνει). Το δεύτερο περιλαμβάνει, για όσα Άρθρα έχουν κείμενο μη χωρισμένο σε Παραγράφους, κάποια από τα μεταδεδομένα των Άρθρων αυτών και το κείμενό τους. Το τρίτο τμήμα περιλαμβάνει, για κάθε Παράγραφο, τα μεταδεδομένα της και το κείμενό της.

Υπάρχει ξεχωριστό τμήμα στο XSLT αρχείο που ελέγχει αν η πηγή νομοθεσίας έχει και Κεφάλαια (χωρισμένα σε Άρθρα που χωρίζονται σε Παραγράφους). Στην περίπτωση αυτή, ο αλγόριθμος που χρησιμοποιείται στη συνέχεια είναι ο ίδιος, περιλαμβάνοντας, στην αρχή, ένα ακόμη κομμάτι που αφορά την εγγραφή των μεταδεδομένων (URI, αριθμός, τύπος (Chapter), τίτλος και λίστα των URI των Άρθρων) κάθε Κεφαλαίου. Τα Κεφάλαια πάντα χωρίζονται σε Άρθρα και δεν έχουν περιεχόμενο κειμένου, οπότε δεν έχει νόημα να ελέγξουμε κάτι τέτοιο, όπως κάνουμε με τα Άρθρα. Έτσι, παράγεται ένα ακόμα τμήμα στο RDF έγγραφο, το οποίο περιλαμβάνει τα μεταδεδομένα που αναφέρθηκαν για κάθε Κεφάλαιο.

Η δομή των παραγόμενων από αυτή τη διαδικασία RDF Σημασιολογικών Εγγράφων έχει περιγραφεί εκτενώς στην Ενότητα 3.2.

5

Σενάρια Χρήσης

Τα RDF σημασιολογικά έγγραφα που δημιουργήσαμε δε θα είχαν κανένα νόημα αν οι πληροφορίες που περιέχουν δεν ήταν διαθέσιμες με κάποιον τρόπο. Η ουσία, εξάλλου, του Σημασιολογικού Ιστού είναι, μεταξύ άλλων, η ευκολότερη πρόσβαση στη γνώση. Για το λόγο αυτό, παρέχουμε πρόσβαση σε έναν διακομιστή (server) τύπου Apache Jena Fuseki, με στόχο τη δημοσίευση των Διασυνδεδεμένων Δεδομένων, αλλά και την υποβολή ερωτημάτων SPARQL σε αυτά. Ο χρήστης που θέλει να θέσει SPARQL ερωτήματα στα δεδομένα μας μπορεί είτε να χρησιμοποιήσει το αντίστοιχο **endpoint**⁶⁴ που ήδη υπήρχε στο διακομιστή και να πληκτρολογήσει τα SPARQL ερωτήματα που θέλει, είτε να χρησιμοποιήσει τη σελίδα «Custom Queries», η οποία παρέχει έξι κουμπιά με έξι διαφορετικά αυτοματοποιημένα SPARQL ερωτήματα. Η σελίδα αυτή είναι επισκέψιμη στον παρακάτω σύνδεσμο:

<http://diplomatikes.dslab.ece.ntua.gr:3039/querypage.html>

5.1 Συγκεντρωτική Λίστα Είδους Νομοθετήματος έως συγκεκριμένη Ημερομηνία

Ας υποθέσουμε πως ένας γλωσσολόγος θέλει να συγκεντρώσει όλα τα Προεδρικά Διατάγματα που δημοσιεύτηκαν έως και το 2013 έτσι ώστε να μελετήσει τις αλλαγές που έγιναν στο ύφος και τη γλώσσα που χρησιμοποιείται σε αυτά με το πέρασμα του χρόνου. Ένα κατάλληλο ερώτημα σε γλώσσα SPARQL για να συγκεντρώσει όλα αυτά τα Προεδρικά Διατάγματα είναι το πρώτο SPARQL ερώτημα (Q1) της σελίδας «Custom Queries»:

64 <http://diplomatikes.dslab.ece.ntua.gr:3039/#/dataset/eligr/query>

```

1 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
2 PREFIX eli: <http://data.europa.eu/eli#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4
5 SELECT ?Description ?id_local ?date_publication
6 WHERE {
7   ?Description eli:id_local ?id_local.
8   ?Description eli:date_publication ?date_publication.
9   ?Description eli:ResourceType ?type.
10  FILTER (year(?date_publication)<=2013 && ?type=<http://legislation.ntua.gr/eli#Presidential_Decree>)
11 }

```

Εικόνα 35: SPARQL Ερώτημα Q1

Τα πρώτα 20 από τα 167 αποτελέσματα που επιστρέφονται από το παραπάνω SPARQL ερώτημα απεικονίζονται στον παρακάτω Πίνακα. Το συγκεκριμένο ερώτημα SPARQL επιστρέφει στο χρήστη το URI, το μοναδικό αναγνωριστικό και την ημερομηνία δημοσίευσης για όλα τα Προεδρικά Διαγράμματα που δημοσιεύτηκαν έως και το 2013. Έχοντας τα URI όλων των Προεδρικών Διαταγμάτων που αναζητά, ο γλωσσολόγος μπορεί να ανατρέχει σε όποιο επιθυμεί από αυτά, έτσι ώστε να μελετήσει το περιεχόμενό του.

URI	Μοναδικό Αναγνωριστικό	Ημερομηνία Δημοσίευσης
http://legislation.ntua.gr/eli/pd/2013/2/	2/ΠΡΩΤΟ/2013-01-09	2013-01-09
http://legislation.ntua.gr/eli/pd/2013/6/	6/ΠΡΩΤΟ/2013-01-21	2013-01-21
http://legislation.ntua.gr/eli/pd/2013/13/	13/ΠΡΩΤΟ/2013-02-14	2013-02-14
http://legislation.ntua.gr/eli/pd/2013/14/	14/ΠΡΩΤΟ/2013-02-14	2013-02-14
http://legislation.ntua.gr/eli/pd/2013/15/	15/ΠΡΩΤΟ/2013-02-14	2013-02-14
http://legislation.ntua.gr/eli/pd/2013/21/	21/ΠΡΩΤΟ/2013-02-26	2013-02-26
http://legislation.ntua.gr/eli/pd/2013/20/	20/ΠΡΩΤΟ/2013-02-26	2013-02-26
http://legislation.ntua.gr/eli/pd/2013/28/	28/ΠΡΩΤΟ/2013-03-12	2013-03-12
http://legislation.ntua.gr/eli/pd/2013/29/	29/ΠΡΩΤΟ/2013-03-15	2013-03-15
http://legislation.ntua.gr/eli/pd/2013/30/	30/ΠΡΩΤΟ/2013-03-15	2013-03-15
http://legislation.ntua.gr/eli/pd/2013/39/	39/ΠΡΩΤΟ/2013-03-20	2013-03-20
http://legislation.ntua.gr/eli/pd/2013/43/	43/ΠΡΩΤΟ/2013-04-05	2013-04-05
http://legislation.ntua.gr/eli/pd/2013/45/	45/ΠΡΩΤΟ/2013-04-16	2013-04-16
http://legislation.ntua.gr/eli/pd/2013/46/	46/ΠΡΩΤΟ/2013-04-18	2013-04-18
http://legislation.ntua.gr/eli/pd/2013/47/	47/ΠΡΩΤΟ/2013-04-18	2013-04-18
http://legislation.ntua.gr/eli/pd/2013/48/	48/ΠΡΩΤΟ/2013-04-18	2013-04-18
http://legislation.ntua.gr/eli/pd/2013/49/	49/ΠΡΩΤΟ/2013-04-22	2013-04-22
http://legislation.ntua.gr/eli/pd/2013/53/	53/ΠΡΩΤΟ/2013-04-26	2013-04-26
http://legislation.ntua.gr/eli/pd/2013/61/	61/ΠΡΩΤΟ/2013-05-14	2013-05-14
http://legislation.ntua.gr/eli/pd/2013/62/	62/ΠΡΩΤΟ/2013-05-14	2013-05-14

Πίνακας 3: Αποτελέσματα SPARQL Ερωτήματος Q1

5.2 Συγκεντρωτική Λίστα Είδους Νομοθετήματος συγκεκριμένου Έτους

Σε αυτό το σενάριο χρήσης, ένας κυβερνητικός σύμβουλος θέλει να συγκεντρώσει και να μελετήσει όλους τους Νόμους που δημοσιεύτηκαν το 2015 έτσι ώστε να σχολιάσει την πολιτική της κυβέρνησης του συγκεκριμένου έτους. Το κατάλληλο SPARQL ερώτημα για αυτήν την περίπτωση βρίσκεται στο δεύτερο κουμπί της σελίδας «Custom Queries» και θέτει στα δεδομένα μας το εξής SPARQL ερώτημα (Q2):


```

1 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
2 PREFIX eli: <http://data.europa.eu/eli#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4
5 SELECT ?Description ?id_local ?date_publication
6 WHERE {
7   ?Description eli:id_local ?id_local.
8   ?Description eli:date_publication ?date_publication.
9   ?Description eli:ResourceType ?type.
10  FILTER (year(?date_publication)=2015 && ?type=<http://legislation.ntua.gr/eli#Act>)
11 }

```

Εικόνα 36: SPARQL Ερώτημα Q2

Παρόμοια με το προηγούμενο, το συγκεκριμένο SPARQL ερώτημα επιστρέφει στον κυβερνητικό σύμβουλο όλους τους Νόμους που δημοσιεύτηκαν το 2015, έτσι ώστε, μέσω των URI τους, να ανατρέξει σε όσους από αυτούς επιθυμεί. Τα αποτελέσματα του SPARQL ερωτήματος αυτού απεικονίζονται στον επόμενο Πίνακα, όπου φαίνεται το URI, το μοναδικό αναγνωριστικό και η ημερομηνία δημοσίευσης των Νόμων που δημοσιεύτηκαν το 2015.

URI	Μοναδικό Αναγνωριστικό	Ημερομηνία Δημοσίευσης
http://legislation.ntua.gr/eli/act/2015/4321/	4321/ΠΡΩΤΟ/2015-03-21	2015-03-21
http://legislation.ntua.gr/eli/act/2015/4323/	4323/ΠΡΩΤΟ/2015-04-27	2015-04-27
http://legislation.ntua.gr/eli/act/2015/4324/	4324/ΠΡΩΤΟ/2015-04-29	2015-04-29
http://legislation.ntua.gr/eli/act/2015/4325/	4325/ΠΡΩΤΟ/2015-05-11	2015-05-11
http://legislation.ntua.gr/eli/act/2015/4326/	4326/ΠΡΩΤΟ/2015-05-13	2015-05-13
http://legislation.ntua.gr/eli/act/2015/4327/	4327/ΠΡΩΤΟ/2015-05-14	2015-05-14
http://legislation.ntua.gr/eli/act/2015/4329/	4329/ΠΡΩΤΟ/2015-06-02	2015-06-02
http://legislation.ntua.gr/eli/act/2015/4330/	4330/ΠΡΩΤΟ/2015-06-16	2015-06-16
http://legislation.ntua.gr/eli/act/2015/4331/	4331/ΠΡΩΤΟ/2015-07-02	2015-07-02
http://legislation.ntua.gr/eli/act/2015/4332/	4332/ΠΡΩΤΟ/2015-07-09	2015-07-09
http://legislation.ntua.gr/eli/act/2015/4333/	4333/ΠΡΩΤΟ/2015-07-11	2015-07-11
http://legislation.ntua.gr/eli/act/2015/4334/	4334/ΠΡΩΤΟ/2015-07-16	2015-07-16
http://legislation.ntua.gr/eli/act/2015/4334/	4335/ΠΡΩΤΟ/2015-07-23	2015-07-23

Πίνακας 4: Αποτελέσματα SPARQL Ερωτήματος Q2

5.3 Συγκεντρωτική Λίστα Άρθρων συγκεκριμένης Πηγής Νομοθεσίας

Έστω ότι ένας πολίτης θέλει να διαβάσει το τμήμα του Προεδρικού Διατάγματος Π.Δ. 1/2015 που αφορά ένα συγκεκριμένο θέμα. Για να μη χρειαστεί να διαβάσει ολόκληρο το Προεδρικό Διάταγμα, θα ήταν χρήσιμο να έχει μια λίστα με όλα τα άρθρα του και τους τίτλους τους, έτσι ώστε να μεταβεί απευθείας στο άρθρο ή τα άρθρα που τον ενδιαφέρουν. Κατάλληλο για την περίπτωση του είναι το SPARQL ερώτημα Q3 της σελίδας «Custom Queries»:

```

1 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
2 PREFIX eli: <http://data.europa.eu/eli#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4
5 SELECT ?partDescription ?number ?title
6 WHERE {
7   ?Description eli:id_local ?id_local.
8   ?Description eli:has_part ?partDescription.
9   ?partDescription eli:number ?number.
10  ?partDescription eli:subdivision_title ?title.
11  FILTER (?id_local="1/ΠΡΩΤΟ/2015-01-09")
12 }

```

Εικόνα 37: SPARQL Ερώτημα Q3

Το αποτέλεσμα που του επιστρέφεται απεικονίζεται παρακάτω. Επιστρέφεται στον πολίτη το URI, ο αριθμός και ο τίτλος κάθε άρθρου της πηγής νομοθεσίας με μοναδικό αναγνωριστικό «1/ΠΡΩΤΟ/2015-01-09», έτσι ώστε ο πολίτης να βρει το Άρθρο που τον ενδιαφέρει μέσω του τίτλου του και να μεταβεί σε αυτό χρησιμοποιώντας το αντίστοιχο URI.

URI	Αριθμός Άρθρου	Τίτλος Άρθρου
http://legislation.ntua.gr/eli/pd/2015/1/article/1/	Άρθρο 1	Πεδίο εφαρμογής
http://legislation.ntua.gr/eli/pd/2015/1/article/2/	Άρθρο 2	Διαδικασία απόσπασης ή μετακίνησης
http://legislation.ntua.gr/eli/pd/2015/1/article/3/	Άρθρο 3	Θέματα υπηρεσιακής και μισθολογικής κατάστασης
http://legislation.ntua.gr/eli/pd/2015/1/article/4/	Άρθρο 4	Αρμοδιότητες - Αξιολόγηση Διάρκεια απόσπασης ή μετακίνησης
http://legislation.ntua.gr/eli/pd/2015/1/article/5/	Άρθρο 5	Μεταβατικές διατάξεις
http://legislation.ntua.gr/eli/pd/2015/1/article/6/	Άρθρο 6	Έναρξη ισχύος

Πίνακας 5: Αποτελέσματα SPARQL Ερωτήματος Q3

5.4 Συγκεντρωτική Λίστα Άρθρων συγκεκριμένης Πηγής Νομοθεσίας και αριθμός των Παραγράφων τους

Υποθέτουμε πως θέλουμε να δημιουργήσουμε έναν πίνακα περιεχομένων του Προεδρικού Διατάγματος Π.Δ. 25/2011, ο οποίος περιλαμβάνει τον αριθμό παραγράφων από τις οποίες αποτελείται κάθε Άρθρο του. Ένα κατάλληλο SPARQL ερώτημα για αυτό το σκοπό είναι το SPARQL ερώτημα Q4 της σελίδας «Custom Queries»:

```

1 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
2 PREFIX eli: <http://data.europa.eu/eli#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4
5 SELECT ?number ?title (count(?par) as ?paragraphs)
6 WHERE {
7   ?Description eli:has_part ?art.
8   ?art eli:subdivision_title ?title.
9   ?art eli:number ?number.
10  OPTIONAL {?art eli:has_part ?par}.
11  ?Description eli:id_local ?id_local.
12  FILTER (?id_local="25/ΠΡΩΤΟ/2011-04-11")
13 }
14 GROUP BY ?number ?title
15 ORDER BY ?number

```

Εικόνα 38: SPARQL Ερώτημα Q4

Το αποτέλεσμα που επιστρέφεται βρίσκεται στον ακόλουθο Πίνακα. Το ερώτημα αυτό επιστρέφει στο χρήστη τον αριθμό, τον τίτλο και τον αριθμό των παραγράφων για κάθε άρθρο της πηγής νομοθεσίας με μοναδικό αναγνωριστικό (id_local) «25/ΠΡΩΤΟ/2011-04-11». Ο χρήστης στη συνέχεια μπορεί να χρησιμοποιήσει τα δεδομένα που του επιστράφηκαν ώστε να δημιουργήσει τον πίνακα περιεχομένων του Π.Δ. 25/2021.

Αριθμός Άρθρου	Τίτλος Άρθρου	Αριθμός Παραγράφων Άρθρου
Άρθρο 1	Ίδρυση - Έδρα	0
Άρθρο 2		0
Άρθρο 3	Καθήκοντα - Αρμοδιότητες οργάνων	2
Άρθρο 4	Έναρξη ισχύος	0

Πίνακας 6: Αποτελέσματα SPARQL Ερωτήματος Q4

5.5 Συγκεντρωτική Λίστα Αναφορών σε άλλες Πηγές Νομοθεσίας συγκεκριμένης Πηγής Νομοθεσίας

Σε αυτό το σενάριο χρήσης, ένας δικηγόρος θέλει, για μια υπόθεσή του, να διαβάσει το Νόμο 4038/2012 και να βρει όλες τις αναφορές που περιέχει έτσι ώστε να τις μελετήσει και αυτές. Κατάλληλο SPARQL ερώτημα για αυτήν την περίπτωση είναι το πέμπτο SPARQL ερώτημα (Q5) της σελίδας «Custom Queries» που απεικονίζεται παρακάτω:

```

1 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
2 PREFIX eli: <http://data.europa.eu/eli#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4
5 SELECT ?citations
6 WHERE {
7   ?Description eli:id_local ?id_local.
8   ?Description eli:cites ?citations.
9   FILTER (?id_local="4038/ΠΡΩΤΟ/2012-02-02")
10 }

```

Εικόνα 39: SPARQL Ερώτημα Q5

Τα αποτελέσματα που επιστρέφει το συγκεκριμένο ερώτημα απεικονίζονται παρακάτω. Επιστρέφονται στο δικηγόρο τα URI (στήλη citations) από όλες τις πηγές νομοθεσίας που υπάρχουν ως αναφορές στην πηγή νομοθεσίας με μοναδικό αναγνωριστικό (id_local) «4038/ΠΡΩΤΟ/2012-02-02», έτσι ώστε να τα χρησιμοποιήσει για την ανάκτηση των νομικών κειμένων που επιθυμεί.

Αναφορές
http://legislation.ntua.gr/eli/pd/2005/81
http://legislation.ntua.gr/eli/pd/2007/118
http://legislation.ntua.gr/eli/pd/1996/50

Πίνακας 7: Αποτελέσματα SPARQL Ερωτήματος Q5

5.6 Συγκεντρωτική Λίστα Πηγών Νομοθεσίας που περιέχουν Αναφορές σε άλλες Πηγές Νομοθεσίας και Αριθμός των Αναφορών αυτών σε Φθίνουσα Σειρά

Ας υποθέσουμε ότι η ΑΑΔΕ θέλει να φτιάξει στην ιστοσελίδα της έναν συγκεντρωτικό πίνακα αναφορών, έναν πίνακα, δηλαδή, ο οποίος περιέχει όλες τις πηγές νομοθεσίας που περιέχουν αναφορές σε άλλες πηγές νομοθεσίας, καθώς και τον αριθμό των αναφορών αυτών. Θέλει, ακόμα, να ταξινομήσει αυτές τις πηγές νομοθεσίας ανάλογα με τον αριθμό αναφορών τους, για να εντοπίσει εκείνες με τις περισσότερες αναφορές έτσι ώστε να αρχίσει να δημιουργεί «συνδέσεις» (hyperlinks) μεταξύ των συσχετιζόμενων νομοθετικών πηγών. Κατάλληλο SPARQL ερώτημα για αυτήν την εργασία είναι το παρακάτω SPARQL ερώτημα Q6 της σελίδας «Custom Queries»:

```

1 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
2 PREFIX eli: <http://data.europa.eu/eli#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4
5 SELECT ?id_local ?title (count(?cit) as ?citations)
6 WHERE {
7   ?Description eli:id_local ?id_local.
8   ?Description eli:title ?title.
9   ?Description eli:cites ?cit.
10 }
11 GROUP BY ?title ?id_local
12 ORDER BY DESC (?citations)

```

Εικόνα 40: SPARQL Ερώτημα Q6

Το ερώτημα αυτό επιστρέφει 163 αποτελέσματα, τα πρώτα 16 εκ των οποίων απεικονίζονται στην επόμενη σελίδα, σε Πίνακα που παράγεται όταν θέσουμε το συγκεκριμένο SPARQL ερώτημα στη σελίδα «Custom Queries». Επιστρέφονται το μοναδικό αναγνωριστικό, ο τίτλος και ο αριθμός αναφορών σε άλλες πηγές νομοθεσίας (στήλη citations) όσων πηγών νομοθεσίας περιέχουν τέτοιες αναφορές. Οι πηγές νομοθεσίας εμφανίζονται κατά φθίνουσα σειρά, από αυτήν που περιέχει τις περισσότερες αναφορές σε άλλες πηγές νομοθεσίας σε αυτήν που περιέχει τις λιγότερες. Μέσω της χρήσης κατάλληλων προγραμμάτων, η ΑΑΔΕ μπορεί να αντλήσει όλα αυτά τα δεδομένα και να δημιουργήσει αυτόματα τον συγκεντρωτικό πίνακα αναφορών που επιθυμεί, έτσι ώστε, στη συνέχεια, να μελετήσει την κάθε πηγή νομοθεσίας του πίνακα αυτού ξεχωριστά ώστε να δημιουργήσει τις απαραίτητες συνδέσεις με τις αναφορές της.

Μοναδικό Αναγνωριστικό	Τίτλος	Αριθμός Αναφορών
4254/ΠΡΩΤΟ/2014-04-07	Μέτρα στήριξης και ανάπτυξης της ελληνικής οικονομίας στο πλαίσιο εφαρμογής του ν. 4046/2012 και άλλες διατάξεις.	91
4072/ΠΡΩΤΟ/2012-04-11	Βελτίωση επιχειρηματικού περιβάλλοντος - Νέα εταιρική μορφή - Σήματα - Μεσίτες Ακινήτων - Ρύθμιση θεμάτων ναυτιλίας, λιμένων και αλιείας και άλλες διατάξεις.	47
3943/ΠΡΩΤΟ/2011-03-31	Καταπολέμηση της φοροδιαφυγής, στελέχωση των ελεγκτικών υπηρεσιών και άλλες διατάξεις αρμοδιότητας Υπουργείου Οικονομικών.	38
3966/ΠΡΩΤΟ/2011-05-24	Θεσμικό πλαίσιο των Πρότυπων Πειραματικών Σχολείων, Ίδρυση Ινστιτούτου Εκπαιδευτικής Πολιτικής, Οργάνωση του Ινστιτούτου Τεχνολογίας Υπολογιστών και Εκδόσεων «ΔΙΟΦΑΝΤΟΣ» και λοιπές διατάξεις.	34
3966/ΠΡΩΤΟ/2011-05-24	Ρυθμίσεις Ηλεκτρονικών Επικοινωνιών, Μεταφορών, Δημοσίων Έργων και άλλες διατάξεις.	30
4141/ΠΡΩΤΟ/2013-04-05	Επενδυτικά εργαλεία ανάπτυξης, παροχή πιστώσεων και άλλες διατάξεις.	30
4075/ΠΡΩΤΟ/2012-04-11	Θέματα Κανονισμού Ασφάλισης ΙΚΑ - ΕΤΑΜ, Ασφαλιστικών Φορέων, προσαρμογή της νομοθεσίας στην Οδηγία 2010/18/ΕΕ και λοιπές διατάξεις.	24
4152/ΠΡΩΤΟ/2013-05-09	Επείγοντα μέτρα εφαρμογής των νόμων 4046/2012, 4093/2012 και 4127/2013.	24
4002/ΠΡΩΤΟ/2011-08-22	Τροποποίηση της συνταξιοδοτικής νομοθεσίας του Δημοσίου - Ρυθμίσεις για την ανάπτυξη και τη δημοσιονομική εξυγίανση - Θέματα αρμοδιότητας Υπουργείων Οικονομικών, Πολιτισμού και Τουρισμού και Εργασίας και Κοινωνικής Ασφάλισης.	22
4071/ΠΡΩΤΟ/2012-04-11	Ρυθμίσεις για την τοπική ανάπτυξη, την αυτοδιοίκηση και την αποκεντρωμένη διοίκηση Ενσωμάτωση Οδηγίας 2009/50/ΕΚ	22
3996/ΠΡΩΤΟ/2011-08-05	Αναμόρφωση του Σώματος Επιθεωρητών Εργασίας, ρυθμίσεις θεμάτων Κοινωνικής Ασφάλισης και άλλες διατάξεις.	21
4030/ΠΡΩΤΟ/2011-11-25	Νέος τρόπος έκδοσης αδειών δόμησης, ελέγχου κατασκευών και λοιπές διατάξεις.	20
4052/ΠΡΩΤΟ/2012-03-01	Νόμος αρμοδιότητας Υπουργείων Υγείας και Κοινωνικής Αλληλεγγύης και Εργασίας και Κοινωνικής Ασφάλισης για εφαρμογή του νόμου «Έγκριση των Σχεδίων Συμβάσεων Χρηματοδοτικής Διευκόλυνσης μεταξύ του Ευρωπαϊκού Ταμείου Χρηματοπιστωτικής Σταθερότητας (Ε.Τ.Χ.Σ.), της Ελληνικής Δημοκρατίας και της Τράπεζας της Ελλάδος, του Σχεδίου του Μνημονίου Συνεννόησης μεταξύ της Ελληνικής Δημοκρατίας, της Ευρωπαϊκής Επιτροπής και της Τράπεζας της Ελλάδος και άλλες επείγουσες διατάξεις για τη μείωση του δημοσίου χρέους και τη διάσωση της εθνικής οικονομίας» και άλλες διατάξεις.	19
3986/ΠΡΩΤΟ/2011-07-01	Επείγοντα Μέτρα Εφαρμογής Μεσοπρόθεσμου Πλαισίου Δημοσιονομικής Στρατηγικής 2012-2015.	17
4058/ΠΡΩΤΟ/2012-03-22	Παροχή υπηρεσιών ασφαλείας από ένοπλους φρουρούς σε εμπορικά πλοία και άλλες διατάξεις.	17
4316/ΠΡΩΤΟ/2014-12-24	Ίδρυση παρατηρητηρίου άνοιας, βελτίωση περιγεννητικής φροντίδας, ρυθμίσεις θεμάτων αρμοδιότητας Υπουργείου Υγείας και άλλες διατάξεις.	17

Πίνακας 8: Αποτελέσματα SPARQL Ερωτήματος Q6

6

Επίλογος

6.1 Σύνοψη

Η παρούσα Διπλωματική Εργασία είχε στόχο την δημοσίευση της Ελληνικής Νομοθεσίας στον Σημασιολογικό Ιστό με τη μορφή Διασυνδεδεμένων Δεδομένων. Για να επιτευχθεί ο στόχος αυτός, χρειάστηκε να μελετηθούν τρία διαφορετικά αντικείμενα: οι βασικές έννοιες και τεχνολογίες που αφορούν τον Σημασιολογικό Ιστό, οι ήδη υπάρχουσες υλοποιήσεις που χρησιμοποιούν τεχνολογίες του Σημασιολογικού Ιστού για τη δημοσίευση νομικών Διασυνδεδεμένων Δεδομένων και οι ιδιαιτερότητες της ελληνικής νομοθεσίας.

Μέσω της μελέτης αυτής, οδηγηθήκαμε στη χρήση του ευρωπαϊκού αναγνωριστικού ELI, το οποίο αποτελεί την νομική οντολογία-πρότυπο που προτείνει η Ευρωπαϊκή Ένωση στα κράτη-μέλη της, έτσι ώστε να επιτευχθεί ο εκσυγχρονισμός της δημοσίευσης των νομικών τους πληροφοριών και να διευκολυνθεί η μεταξύ τους συνεργασία. Αφού εμπλουτίσαμε την οντολογία ELI έτσι ώστε να καλύπτονται οι ανάγκες και ιδιαιτερότητες του ελληνικού νομικού συστήματος, δημιουργήσαμε μια επέκτασή της, το ELIgr.

Στη συνέχεια, προχωρήσαμε στη δημιουργία RDF νομικών σημασιολογικών εγγράφων που ακολουθούν το πρότυπο ELIgr. Για να γίνει αυτό, είχαμε στη διάθεσή μας τμήμα της ελληνικής νομοθεσίας σε μορφή σημασιολογικών εγγράφων με βάση το πρότυπο LegalDocML. Αφού η τελική μορφή των RDF αρχείων οριστικοποιήθηκε, η διαδικασία μετατροπής των LegalDocML αρχείων σε σημασιολογικά έγγραφα RDF αυτοματοποιήθηκε μέσω της δημιουργίας και χρήσης ενός XSLT stylesheet, το οποίο παρέχει στον υπολογιστή τους «κανόνες» με βάση τους οποίους γίνεται η μετατροπή αυτή.

Τα RDF νομικά σημασιολογικά έγγραφα που προέκυψαν από την παραπάνω μετατροπή τα είναι διαθέσιμα ως συλλογή δεδομένων με όνομα eligr σε διακομιστή τύπου Apache Jena Fuseki. Στο SPARQL endpoint του διακομιστή αυτού, ο χρήστης μπορεί να διατυπώσει SPARQL ερωτήματα στην eligr και να του επιστραφούν από το διακομιστή οι πληροφορίες που θέλει. Μέσω του endpoint αυτού, και δημιουργώντας έξι δικά μας SPARQL ερωτήματα, παρουσιάσαμε έξι σενάρια χρήσης του διακομιστή.

Τέλος, μέσω της σελίδας με όνομα «SPARQL Queries» που προσθέσαμε στον διακομιστή, προσφέρεται στο χρήστη η δυνατότητα να θέσει τα ερωτήματα αυτά στη συλλογή δεδομένων eligr χωρίς να χρειάζεται να τα πληκτρολογήσει κάθε φορά. Με αυτόν τον τρόπο, τα παραπάνω σενάρια χρήσης μπορούν να παρουσιάζονται ευκολότερα και γρηγορότερα.

6.2 Συμπεράσματα και Παρατηρήσεις

Κατά τη διάρκεια εκπόνησης της παρούσας διπλωματικής εργασίας έγινε αντιληπτή η επιτακτική ανάγκη για τον εκσυγχρονισμό της δημοσίευσης των νομικών πληροφοριών παγκοσμίως, αλλά και ιδιαίτερος στην Ελλάδα. Υπάρχει μεγάλη ποσότητα Ελληνικών νομικών δεδομένων στο διαδίκτυο, η οποία ωστόσο δεν είναι ούτε σωστά οργανωμένη, ούτε παρέχει στο χρήστη άμεσα την πληροφορία που αναζητά.

Το πρώτο πρόβλημα που συναντά κανείς είναι οι πηγές της νομικής πληροφορίας. Υπάρχουν πολλές ιστοσελίδες, άλλες δημόσιες και άλλες ιδιωτικές, οι οποίες παρέχουν την ελληνική νομοθεσία χωρίς, όμως, να είναι πάντα πλήρεις. Αυτό έχει ως αποτέλεσμα διαφορετικά τμήματα της ελληνικής νομοθεσίας να βρίσκονται σε διαφορετικές ιστοσελίδες, δυσχεραίνοντας, έτσι, την εμπειρία των χρηστών. Επιπλέον, στις περισσότερες από αυτές τις ιστοσελίδες, η αναζήτηση πληροφορίας μπορεί να γίνει μόνο μέσω του ΦΕΚ ή της ημερομηνίας ψήφισης της εκάστοτε πηγής νομοθεσίας, γεγονός που απαιτεί από τον χρήστη να γνωρίζει εκ των προτέρων τις πληροφορίες αυτές όταν αναζητεί τη νομοθεσία πάνω σε κάποιο θέμα, πράγμα που τις περισσότερες φορές δε συμβαίνει.

Ένα ακόμα πρόβλημα είναι η μορφή στην οποία παρέχεται η ελληνική νομοθεσία στο διαδίκτυο. Συνήθως παρέχεται σε κλειστές μορφές αρχείων, όπως αρχεία PDF, ενώ σε ελάχιστες περιπτώσεις συναντάμε html μορφή. Συχνά, μάλιστα, και ιδιαίτερα για παλαιότερες πηγές νομοθεσίας, τα αρχεία PDF αποτελούνται από εικόνες που προκύπτουν από τη σάρωση των πρωτότυπων των πηγών αυτών, με αποτέλεσμα να μην είναι δυνατή η αναζήτηση λέξεων-κλειδιών από τους χρήστες. Σημασιολογικά δεδομένα συναντούμε σπάνια και είτε βρίσκονται σε ιδιωτικές ιστοσελίδες που συχνά απαιτούν πληρωμή, είτε βρίσκονται σε ιστοσελίδες που αφορούν πανεπιστημιακές εργασίες και σταματούν να είναι λειτουργικές μετά από κάποιο διάστημα, ενώ σε καμία από τις δύο αυτές περιπτώσεις η νομοθεσία δεν είναι πλήρης.

Τέλος, τις περισσότερες φορές δεν είναι σαφές αν μια πηγή νομοθεσίας έχει τροποποιηθεί ή καταργηθεί από κάποια μεταγενέστερη. Αυτό στην πράξη έχει ως αποτέλεσμα ο χρήστης να μη γνωρίζει με ασφάλεια αν η πληροφορία που παίρνει είναι έγκυρη ή όχι. Αν συνυπολογίσουμε και το γεγονός ότι τα νομοθετικά κείμενα είναι γραμμένα με υπερβολική τυπικότητα, λόγια γλώσσα και ασάφειες και περιλαμβάνουν πολλές περιττές πληροφορίες και επαναλήψεις, συνειδητοποιούμε πως ο απλός χρήστης είναι πιθανότερο να μη βρει τη σωστή απάντηση στο θέμα για το οποίο ενδιαφέρεται.

Είναι προφανές ότι χρειάζεται να γίνουν ριζικές αλλαγές όχι μόνο στις διαδικασίες δημοσίευσης της ελληνικής νομοθεσίας, αλλά και στις διαδικασίες που αφορούν τη δημιουργία της. Μόνο έτσι θα μπορέσει το Ελληνικό κράτος να παρέχει ουσιαστικές και ποιοτικές υπηρεσίες ενημέρωσης στους πολίτες σχετικά με την ισχύουσα νομοθεσία, ενδυναμώνοντάς τους και εξοπλίζοντάς τους με τη γνώση που χρειάζονται ώστε να είναι περισσότερο ενεργοί, να παίρνουν σωστότερες αποφάσεις για τη ζωή τους και να συμβάλλουν, έτσι, στην κοινωνική αλλά και την ατομική τους εξέλιξη.

6.3 Συνεισφορά και Προτάσεις για το Μέλλον

Με την παρούσα εργασία επιτεύχθηκαν δύο στόχοι. Ο πρώτος και σημαντικότερος στόχος είναι η

δημιουργία της επέκτασης ELIgr του ευρωπαϊκού προτύπου ELI, η οποία είναι προσαρμοσμένη στα ελληνικά νομικά δεδομένα. Δημιουργώντας την επέκταση αυτή, καθορίσαμε τη μορφή των ελληνικών URIs, τη δομή που οφείλουν να ακολουθούν τα RDF σημασιολογικά έγγραφα που περιγράφουν τις ελληνικές πηγές νομοθεσίας και τις πληροφορίες που πρέπει να περιέχουν. Ουσιαστικά, προτείναμε τη «βάση» πάνω στην οποία θα αναπτυχθεί ένα νέο, εκσυγχρονισμένο σύστημα δημοσίευσης Διασυνδεδεμένων νομικών Δεδομένων στην Ελλάδα, ένα σύστημα που θα «μεταφέρει» την ελληνική νομοθεσία από τον Παγκόσμιο στον Σημασιολογικό Ιστό.

Ο δεύτερος στόχος είναι η μετατροπή των ήδη υπάρχοντων σημασιολογικών εγγράφων που αφορούν την ελληνική νομοθεσία σε σημασιολογικά έγγραφα πιο σύγχρονης μορφής. Τα νομικά σημασιολογικά έγγραφα που υπήρχαν ακολουθούν το πρότυπο LegalDocML και, μέσω του XSLT stylesheet που δημιουργήσαμε, αυτοματοποιήσαμε τη διαδικασία μετατροπής τους σε σημασιολογικά έγγραφα RDF που ακολουθούν το πρότυπο ELI. Το πρότυπο ELI είναι ιδιαίτερα σημαντικό διότι, πέρα από το γεγονός πως είναι πιο σύγχρονο, είναι και το πρότυπο που προτείνεται από την Ευρωπαϊκή Ένωση, κάτι που σημαίνει πως η χρήση του θα διευκολύνει την διαλειτουργικότητα των συστημάτων και την επικοινωνία και ανταλλαγή νομικών πληροφοριών με τα υπόλοιπα ευρωπαϊκά κράτη.

Επιπλέον, με τον διακομιστή που εγκαταστήσαμε, και μέσω των SPARQL ερωτημάτων που προσθέσαμε σε αυτόν, είναι δυνατόν να επιτευχθεί ο έλεγχος ορθότητας των σημασιολογικών εγγράφων που δημιουργήσαμε και αυτών που θα δημιουργηθούν στο μέλλον.

Ωστόσο, είναι προφανές πως η παρούσα εργασία αποτελεί μόνο την αρχή της διαδικασίας εκσυγχρονισμού του ελληνικού συστήματος δημοσίευσης νομικών δεδομένων στο διαδίκτυο. Υπάρχει μια σειρά από σημαντικές εργασίες που οφείλουν να γίνουν έτσι ώστε οι ήδη υπάρχουσες νομικές πληροφορίες να αποκτήσουν μορφή Διασυνδεδεμένων Δεδομένων, όπως επίσης και μια σειρά από επεκτάσεις που μπορούν να γίνουν πάνω στο σύστημα που αναπτύξαμε στην παρούσα εργασία. Οι πιο σημαντικές είναι οι εξής:

- Συγκέντρωση όλων των ελληνικών νομικών κειμένων: Αυτή τη στιγμή, τα ελληνικά νομικά κείμενα βρίσκονται στην ιστοσελίδα του Εθνικού Τυπογραφείου και σε άλλες ιστοσελίδες σε ποικίλες μορφές αρχείων (PDF, html κ.λ.π.). Τα κείμενα αυτά πρέπει να συγκεντρωθούν σε μια βάση δεδομένων και να ταξινομηθούν ανάλογα με τη μορφή τους, έτσι ώστε να μπορεί να αρχίσει η διαδικασία μετατροπής τους σε RDF σημασιολογικά έγγραφα.
- Μετατροπή των ήδη υπάρχοντων νομικών κειμένων σε RDF σημασιολογικά έγγραφα: Όπως ήδη αναφέρθηκε, τα ελληνικά νομικά κείμενα αυτή τη στιγμή διατίθενται στο διαδίκτυο σε μια ποικιλία μορφών (PDF, docx, html κ.λ.π.). Για κάθε μια από τις μορφές αυτές πρέπει να χρησιμοποιηθούν τα αντίστοιχα και κατάλληλα εργαλεία εξαγωγής και ανάλυσης κειμένου ώστε η μετατροπή τους σε σημασιολογικά έγγραφα RDF να αυτοματοποιηθεί όσο το δυνατόν περισσότερο. Ειδικά για τα αρχεία PDF μορφής που αποτελούνται από σαρωμένες εικόνες και όχι από κείμενο, είναι απαραίτητη και η χρήση εργαλείων οπτικής αναγνώρισης χαρακτήρων (Optical Character Recognition - OCR). Είναι σίγουρο ότι θα χρειαστεί να γίνουν και χειροκίνητες παρεμβάσεις, δεδομένου ότι τα ελληνικά νομικά κείμενα έχουν ασυνέπειες στη δομή τους και ότι τα παραπάνω εργαλεία δεν είναι πάντα 100% ακριβή. Ωστόσο, η συνεχής και ταχεία εξέλιξη των μοντέλων και εργαλείων τεχνητής νοημοσύνης (Artificial Intelligence - AI) είναι βέβαιο πως μπορεί να συμβάλλει στη μείωση των χειροκίνητων παρεμβάσεων που θα χρειαστούν.
- Διάθεση των νομικών Διασυνδεδεμένων Δεδομένων στο ευρύ κοινό: Όταν η μετατροπή της νομικής πληροφορίας σε σημασιολογικά δεδομένα ολοκληρωθεί, είναι απαραίτητο να δημιουργηθεί μια πλατφόρμα πρόσβασης που θα διαθέτει τα δεδομένα αυτά στο ευρύ κοινό. Η πλατφόρμα αυτή θα πρέπει να είναι απλή στη χρήση, να παρέχει ποικίλους τρόπους και κριτήρια αναζήτησης νομικών πληροφοριών στο χρήστη και να εκμεταλλεύεται τις δυνατότητες των Διασυνδεδεμένων Δεδομένων έτσι ώστε η παρουσίαση της πληροφορίας να είναι τέτοια, ώστε η πληροφορία να γίνεται εύκολα κατανοητή. Σε επόμενο στάδιο, η πλατφόρμα αυτή θα μπορούσε να συνδεθεί και με άλλες πλατφόρμες κρατών της Ευρώπης που χρησιμοποιούν το Σημασιολογικό Ιστό και το πρότυπο ELI για την οργάνωση των νομικών

Διασυνδεδεμένων Δεδομένων τους.

- Χρήση και έλεγχος του ELIgr για τα υπόλοιπα είδη πηγών νομοθεσίας: Αυτή τη στιγμή, η επέκταση ELIgr έχει χρησιμοποιηθεί και ελεγχθεί μόνο για νομικά κείμενα που αφορούν Νόμους και Προεδρικά Διατάγματα, ενώ έχουν μοντελοποιηθεί και τέσσερις Εγκύκλιοι και Αποφάσεις. Είναι σημαντικό να μοντελοποιηθούν και άλλου είδους πηγές νομοθεσίας (όπως Σύνταγμα, Διεθνείς Συμβάσεις κ.λ.π.), έτσι ώστε να αναγνωριστούν οι ιδιαιτερότητες της καθεμίας και να καθοριστεί η μορφή και η δομή των RDF σημασιολογικών εγγράφων για κάθε περίπτωση. Αντίστοιχα, πρέπει στον διακομιστή να προστεθούν νέα SPARQL ερωτήματα στη σελίδα «SPARQL Queries», έτσι ώστε να ελέγχεται ευκολότερα η ορθότητα των νέων δεδομένων που θα προστεθούν.
- Χρήση εξειδικευμένων εργαλείων ανάλυσης κειμένου για εξαγωγή περισσότερων και ποιοτικότερων πληροφοριών και μεταδεδομένων: Τα σημασιολογικά έγγραφα LegalDocML που χρησιμοποιήσαμε για την παρούσα εργασία περιέχουν αρκετά μεταδεδομένα. Ωστόσο, με ανάγνωση του περιεχομένου του νομικού κειμένου βλέπουμε πως υπάρχουν επιπλέον μεταδεδομένα που δεν έχουν αναγνωριστεί ως τέτοια. Τέτοια μεταδεδομένα αποτελούν το ΦΕΚ της πηγής νομοθεσίας, η αναφοράς σε άλλες πηγές νομοθεσίας ή και σε άλλα κείμενα, το κατά πόσο η πηγή νομοθεσίας αποτελεί τροποποίηση κάποιας άλλης, το κατά πόσο η τροποποίηση είναι πρόσθεση, συνένωση, επαναρίθμηση, διαχωρισμός ή αντικατάσταση της προηγούμενης πηγής νομοθεσίας, τα πρόσωπα που υπέγραψαν την πηγή νομοθεσίας κ.λ.π.. Σε άλλες περιπτώσεις, λόγω κακής ανάλυσης κειμένου, υπάρχουν λάθη, όπως η σύμπτυξη τίτλων άρθρων με το περιεχόμενό τους. Το XSLT stylesheet δεν είναι εργαλείο ανάλυσης κειμένου, γεγονός που σημαίνει ότι τα παραγόμενα RDF σημασιολογικά έγγραφα του συστήματός μας «κληρονομούν» τα λάθη και τις παραλείψεις που υπάρχουν στα μεταδεδομένα των LegalDocML αρχείων. Για την επίλυση του προβλήματος αυτού είναι απαραίτητη η χρήση κατάλληλου λογισμικού ανάλυσης κειμένου, με την παράλληλη συνδρομή μοντέλων τεχνητής νοημοσύνης και χειροκίνητης εργασίας όπου αυτό κριθεί απαραίτητο.
- Δημοσίευση Διασυνδεδεμένων Δεδομένων που αφορά τα πολιτικά πρόσωπα της χώρας: Στο παρόν στάδιο, τα RDF σημασιολογικά κείμενα που έχουμε παράξει αναφέρουν μόνο την ιδιότητα των προσώπων που τα υπέγραψαν (Υπουργός, Πρόεδρος της Δημοκρατίας κ.λ.π.). Θα ήταν χρήσιμο να αναπτυχθεί μια οντολογία η οποία θα περιγράφει τα πολιτικά πρόσωπα της χώρας, έτσι ώστε να μπορούμε να εμπλουτίσουμε κάθε νομικό κείμενο με μεταδεδομένα που αφορούν το ποιος το ενέκρινε ή το υπέγραψε, τι πολιτική θέση κατείχε κ.λ.π..
- Ταξινόμηση με βάση τη θεματολογία: Σημαντική αναβάθμιση θα αποτελέσει η ταξινόμηση των νομικών κειμένων με βάση τη θεματολογία στην οποία αναφέρονται, καθώς ο χρήστης συνήθως ψάχνει τη νομοθεσία πάνω σε ένα συγκεκριμένο θέμα χωρίς να γνωρίζει το ΦΕΚ ή την ημερομηνία ψήφισης των νομικών κειμένων που αναφέρονται σε αυτό. Η ταξινόμηση αυτή μπορεί να γίνει με την πρόσθεση κατάλληλων λέξεων-κλειδιών ως μεταδεδομένα κάθε πηγής νομοθεσίας. Η εύρεση των λέξεων αυτών είναι χρονοβόρα, καθώς απαιτεί μελέτη του κάθε νομικού κειμένου ξεχωριστά, ωστόσο μπορεί να βοηθηθεί από κατάλληλους αναλυτές κειμένων και από την εκμετάλλευση των συνεχώς αυξανόμενων δυνατοτήτων της τεχνητής νοημοσύνης.

7

Βιβλιογραφία

-
- [1] Francart, Thomas, et al. "The European legislation identifier." *Knowledge of the Law in the Big Data Age* 317 (2019): 137-148. DOI: [10.3233/FAIA190016](https://doi.org/10.3233/FAIA190016)
 - [2] Σωτηρόπουλος, Δ. Α., and Λ. Χριστόπουλος. "Πολυνομία, Κακονομία και Γραφειοκρατία στην Ελλάδα." *Διανέοσις*, Αθήνα (2016).
 - [3] Hyvönen, Eero, et al. "Linked Data Finland: A 7-star model and platform for publishing and re-using linked datasets." *The Semantic Web: ESWC 2014 Satellite Events: ESWC 2014 Satellite Events*, Anissaras, Crete, Greece, May 25-29, 2014, Revised Selected Papers 11. Springer International Publishing, 2014. DOI: [10.1007/978-3-319-11955-7_24](https://doi.org/10.1007/978-3-319-11955-7_24)
 - [4] Hyvönen, Eero, et al. "Publishing and using legislation and case law as linked open data on the semantic web." *European Semantic Web Conference*. Cham: Springer International Publishing, 2020. DOI: [10.1007/978-3-030-62327-2_19](https://doi.org/10.1007/978-3-030-62327-2_19)
 - [5] Hyvönen, Eero, et al. "LawSampo: A semantic portal on a linked open data service for Finnish legislation and case law." *Proceedings of ESWC*. 2020.
 - [6] Hyvönen, Eero. "Cultural heritage linked data on the semantic web: Three case studies using the sampo model." *VIII Encounter of documentation centres of contemporary art: open linked data and integral management of information in cultural centres*. Artium, Vitoria-Gasteiz, Spain, October (2016): 19-20.

-
- [7] Rodríguez-Doncel, Víctor, et al. "Spanish Legislation as Linked Data." *TERECOM@ JURIX*. 2018.
- [8] Filtz, Erwin, Sabrina Kirrane, and Axel Polleres. "The linked legal data landscape: linking legal data across different countries." *Artificial Intelligence and Law* 29.4 (2021): 485-539. DOI: [10.1007/s10506-021-09282-8](https://doi.org/10.1007/s10506-021-09282-8)
- [9] Avgerinos Loutsaris, Michalis, et al. "Semantic Interoperability for Legal Information: Mapping the European Legislation Identifier (ELI) and Akoma Ntoso (AKN) Ontologies." *Proceedings of the 16th International Conference on Theory and Practice of Electronic Governance*. 2023. DOI: [10.1145/3614321.3614327](https://doi.org/10.1145/3614321.3614327)
- [10] Galiotou, Eleni, and Pavlina Fragkou. "Applying linked data technologies to Greek open government data: a case study." *Procedia-social and behavioral sciences* 73 (2013): 479-486. DOI: [10.1016/j.sbspro.2013.02.080](https://doi.org/10.1016/j.sbspro.2013.02.080)
- [11] Koniaris, Marios, George Papastefanatos, and Yannis Vassiliou. "Towards automatic structuring and semantic indexing of legal documents." *Proceedings of the 20th Pan-Hellenic Conference on Informatics*. 2016. DOI: [10.1145/3003733.3003801](https://doi.org/10.1145/3003733.3003801)
- [12] Koniaris, Marios, George Papastefanatos, and Ioannis Anagnostopoulos. "Solon: A holistic approach for modelling, managing and mining legal sources." *Algorithms* 11.12 (2018): 196. DOI: [10.3390/a11120196](https://doi.org/10.3390/a11120196)
- [13] Chalkidis, Ilias, et al. "Modeling and querying greek legislation using semantic web technologies." *The Semantic Web: 14th International Conference, ESWC 2017, Portorož, Slovenia, May 28–June 1, 2017, Proceedings, Part I* 14. Springer International Publishing, 2017. DOI: [10.1007/978-3-319-58068-5_36](https://doi.org/10.1007/978-3-319-58068-5_36)
- [14] Angelidis, Iosif, et al. "Nomothesia: a linked data platform for Greek legislation." (2018).
- [15] Angelidis, Iosif, Ilias Chalkidis, and Manolis Koubarakis. "Named Entity Recognition, Linking and Generation for Greek Legislation." *JURIX*. 2018.
- [16] Berners-Lee, Tim, James Hendler, and Ora Lassila. "The semantic web." *Scientific american* 284.5 (2001): 34-43.
- [17] Bizer, Christian, Tom Heath, and Tim Berners-Lee. "Linked data: The story so far." *Semantic services, interoperability and web applications: emerging concepts*. IGI global, 2011. 205-227. DOI: [10.4018/jswis.2009081901](https://doi.org/10.4018/jswis.2009081901)
- [18] Horrocks, Ian. "Ontologies and the semantic web." *Communications of the ACM* 51.12 (2008): 58-67. DOI: [10.1145/1409360.1409377](https://doi.org/10.1145/1409360.1409377)
- [19] Wittenburg, Peter, and Daan Broeder. "Metadata overview and the semantic web." *The 3rd International Conference on Language Resources and Evaluation (LREC 2002). Workshop on Tools and Resources in Field Linguistics*. European Language Resources Association, 2002.
- [20] Pon, Raymond K., and David J. Buttlar. *Metadata registry, iso/iec 11179*. No. LLNL-JRNL-400269. Lawrence Livermore National Lab.(LLNL), Livermore, CA (United States), 2008. DOI: [10.1007/978-1-4614-8265-9_907](https://doi.org/10.1007/978-1-4614-8265-9_907)
- [21] Decker, Stefan, Prasenjit Mitra, and Sergey Melnik. "Framework for the semantic Web: an RDF tutorial." *IEEE Internet Computing* 4.6 (2000): 68-73. DOI: [10.1109/4236.895018](https://doi.org/10.1109/4236.895018)
- [22] Grlický, Vladimír. "An overview of RDF model representation formats." *IIT. SRC 2005: Student Research Conference*. 2005.
- [23] Beckett, David, et al. "RDF 1.1 Turtle." *World Wide Web Consortium* (2014): 18-31.
- [24] Berners-Lee, Tim, Dan Connolly, and Sandro Hawke. "Semantic web tutorial using n3." *Twelfth International World Wide Web Conference*. 2003.
- [25] Beckett, Dave, and Brian McBride. "RDF/XML syntax specification (revised)." *W3C recommendation* 10.2.3 (2004).
- [26] Sporny, Manu, et al. "JSON-LD 1.1." *W3C Recommendation*, Jul (2020).

-
- [27] Berners-Lee, Tim, Roy Fielding, and Larry Masinter. Uniform resource identifier (URI): Generic syntax. No. Rfc3986. 2005. DOI: [10.17487/rfc3986](https://doi.org/10.17487/rfc3986)
- [28] Gómez-Pérez, Asunción, and Oscar Corcho. "Ontology languages for the semantic web." *IEEE Intelligent systems* 17.1 (2002): 54-60.
- [29] Nejd, Wolfgang, et al. "The RDF schema specification revisited." In *Modelle und Modellierungssprachen in Informatik und Wirtschaftsinformatik, Modellierung 2000* (2000).
- [30] Horrocks, Ian, Peter F. Patel-Schneider, and Frank Van Harmelen. "Reviewing the design of DAML+ OIL: An ontology language for the semantic web." *AAAI/IAAI 2002* (2002): 792-797.
- [31] McGuinness, Deborah L., and Frank Van Harmelen. "OWL web ontology language overview." *W3C recommendation* 10.10 (2004): 2004.
- [32] World Wide Web Consortium. "OWL 2 web ontology language document overview." (2012). URL (τελευταία επίσκεψη 22/02/2024): <https://www.w3.org/TR/owl2-overview/>
- [33] Harold, Elliotte Rusty, and W. Scott Means. *XML in a nutshell: a desktop quick reference*. " O'Reilly Media, Inc.", 2004.
- [34] World Wide Web Consortium. "Namespaces in XML 1.1." (2006).
- [35] Tidwell, Doug. *XSLT: mastering XML transformations*. " O'Reilly Media, Inc.", 2008.
- [36] Clark, James, and Steve DeRose. "XML path language (XPath)." (1999).
- [37] Walmsley, Priscilla. *XQuery*. " O'Reilly Media, Inc.", 2007.
- [38] DuCharme, Bob. *Learning SPARQL: querying and updating with SPARQL 1.1*. " O'Reilly Media, Inc.", 2013.
- [39] Karvounarakis, Gregory, et al. "RQL: a declarative query language for RDF." *Proceedings of the 11th international conference on World Wide Web*. 2002.
- [40] World Wide Web Consortium. "Best practices for publishing linked data." (2014).
- [41] Bizer, Christian, and Richard Cyganiak. "D2r server-publishing relational databases on the semantic web." *Poster at the 5th international semantic web conference*. Vol. 175. 2006.
- [42] Adida, Ben, et al. "RDFa Core 1.1." 2015-03-17[2019-8-25]. <https://www.w3.org/TR/2015/REC-rdfa-core-20150317> (2007).
- [43] Khare, Rohit, and Tantek Çelik. "Microformats: a pragmatic path to the semantic web." *Proceedings of the 15th international conference on World Wide Web*. 2006.
- [44] Pareti, Paolo, and George Konstantinidis. "A review of shacl: From data validation to schema reasoning for rdf graphs." *Reasoning Web International Summer School* (2021): 115-144. DOI: [10.1007/978-3-030-95481-9_6](https://doi.org/10.1007/978-3-030-95481-9_6)
- [45] Musen, Mark A. "The protégé project: a look back and a look forward." *AI matters* 1.4 (2015): 4-12. DOI: [10.1145/2757001.2757003](https://doi.org/10.1145/2757001.2757003)
- [46] Haase, Peter, et al. "The neon ontology engineering toolkit." *WWW* (2008). URL (τελευταία επίσκεψη 22/02/2024): <https://dl.wqtxts1xzle7.cloudfront.net/30706200/10.1.1.141.4163-libre.pdf>
- [47] Glimm, Birte, et al. "HermiT: an OWL 2 reasoner." *Journal of automated reasoning* 53 (2014): 245-269. DOI: [10.1007/s10817-014-9305-1](https://doi.org/10.1007/s10817-014-9305-1)
- [48] Sirin, Evren, et al. "Pellet: A practical owl-dl reasoner." *Journal of Web Semantics* 5.2 (2007): 51-53. DOI: [10.1016/j.websem.2007.03.004](https://doi.org/10.1016/j.websem.2007.03.004)
- [49] Tsarkov, Dmitry, and Ian Horrocks. "FaCT++ description logic reasoner: System description." *International joint conference on automated reasoning*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006.
- [50] Francart, Thomas, et al. "The European legislation identifier." *Knowledge of the Law in the Big Data Age* 317 (2019): 137-148. DOI: [10.3233/FAIA190016](https://doi.org/10.3233/FAIA190016)
- [51] Taskforce, E. L. I. "ELI technical guide." (2018). URL (τελευταία επίσκεψη 22/02/2024): <https://policycommons.net/artifacts/263228/eli-technical-guide/1047431/>
- [52] Koniaris, Marios. "Οργάνωση Διαχείριση και Ανάκτηση Φ.Ε.Κ. χρησιμοποιώντας Τεχνολογίες

του Σημασιολογικού Ιστού" (2012). URL (τελευταία επίσκεψη 11/03/2024):
<https://apothesis.eap.gr/archive/item/85560>